



UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS – PGLetras

LAYANE KESSIA PEREIRA SOUSA

**“FULANO” TAMBÉM INDETERMINA SUJEITO: FOTOGRAFIA DAS
FORMAS NOMINAIS DE REFERENCIAÇÃO DO SUJEITO NO PORTUGUÊS
FALADO NAS CAPITAIS DO NORDESTE**

São Luís, MA

2022

LAYANE KESSIA PEREIRA SOUSA

**FULANO” TAMBÉM INDETERMINA SUJEITO: FOTOGRAFIA DAS
FORMAS NOMINAIS DE REFERENCIAÇÃO DO SUJEITO NO PORTUGUÊS
FALADO NAS CAPITAIS DO NORDESTE**

Texto apresentado para defesa de dissertação de Mestrado em Letras do Programa de Pós-Graduação em Letras da Universidade Federal do Maranhão – PGLetras UFMA.

Área de concentração: Descrição e Análise do Português Brasileiro.

Orientadora: Profa. Dra. Cibelle Corrêa Béliche Alves.

São Luís, MA

2022

**FULANO” TAMBÉM INDETERMINA SUJEITO: fotografia das formas
nominais de referenciação do sujeito no português falado nas capitais do Nordeste**

BANCA EXAMINADORA

Prof^ª. Dr^ª. Cibelle Corrêa Béliche Alves (UFMA)
Orientadora

Prof^ª. Dr^ª. Carolina Queiroz Andrade (UniCeub, PPGL/UnB)
Membro Externo

Prof. Dr. Wendel Silva dos Santos (UFMA)
Membro Interno

Prof^ª. Dr^ª. Greize Alves da Silva (UFT)
Suplente

São Luís-MA

2022

A Deus, por ser a minha luz quando tudo parecia um campo escuro.

*In memoriam ao meu pai, Paulo Roberto,
com eterno amor e lembrança dedico esta conquista.*

AGRADECIMENTOS

*“Não te ordenei eu? Esforça-te e SÊ VALENTE! Não temas nem desanimes, porque o Senhor teu Deus
estará contigo onde quer que vás”.*

Josué 1:9

Ao longo desta etapa da vida tive o prazer de conviver com pessoas que me incentivaram e apoiaram nesta árdua caminhada. Hoje, expresso a minha gratidão a todos aqueles que contribuíram direta e indiretamente. Sem dúvidas, não poderia deixar de dedicar este trabalho a vocês. Portanto, agradeço:

A Deus, por sempre estar ao meu lado, sendo meu melhor amigo e fiel companheiro. Agradeço por sua bondade e misericórdia e por ser meu porto seguro nos momentos mais difíceis. Graças dou a Ti por se revelar de forma majestosa em cada decisão da minha vida, na certeza de “Porque dele, e por ele, e para ele são todas as coisas; glória, pois, a ele eternamente” Rm 11:36.

A meu pai, Paulo Roberto, a quem tive o prazer de aprender as mais valiosas das lições que guardo em meu coração como quem guarda a mais rara das joias, que, infelizmente, não pôde compartilhar comigo esta realização, mas que sempre vibrava com felicidade por cada conquista recebida. Espero que esta notícia consiga ir além dos limites do concreto, e possa chegar a seus ouvidos. Sigo tentando deixá-lo orgulhoso.

Às minhas irmãs, Luciana e Luana, pelas conversas, conselhos, cuidados, incentivos, dedicação e amor. Sou grata a Deus por tê-las como irmãs/amigas e por acreditarem sempre no meu potencial.

Aos meus sobrinhos, Cecília e Diego, como carinhosamente tenho orgulho de chamar “Filhos do coração”. Agradeço por serem crianças iluminadas que cuidam de mim, assim como eu cuido de vocês. Obrigada por cada afago recebido!

Ao meu querido e amado noivo, Mateus Carvalho, pelo amor, cuidado, preocupação, apoio, orações e por honrar com a promessa de estar sempre ao meu lado. Agradeço meu bem, por estar presente tornando esta jornada mais leve embalada por palavras sábias que me fizeram refletir sobre a vida.

A minha querida orientadora professora Cibelle Béliche, pela orientação sempre estimulante, encorajando em mim a busca por novos conhecimentos. Obrigada por ter sido além de uma orientadora, mas uma amiga que sempre me aconselhava quando precisava. Seu apoio foi fundamental para o bom desenvolvimento deste trabalho.

As minhas amigas Adrienne e Mylena, pelos raros, porém memoráveis encontros que nos ajudaram esquecer um pouco o mundo acadêmico. Gratidão pelas conversas, risadas e parceria. O apoio de vocês foi ímpar.

Aos meus amigos Gabriel e Andreza pelo companheirismo e incentivo. Agradeço pelas conversas icônicas embaladas por risadas memoráveis.

A minha tia Vitória e as minhas primas Patrícia e Ana, que sempre vibraram pelas minhas conquistas.

Aos professores Conceição e Mendes e aos demais amigos/colegas que fazem parte do Projeto ALiMA.

A todos os professores do PG-Letras, pelos ensinamentos transmitidos e conhecimentos compartilhados.

À FAPEMA, pela bolsa de estudo que financiou esta pesquisa e tornou possível a concretização deste trabalho.

Agradeço a todos que contribuíram direta e indiretamente para mais esta etapa da minha vida.

Só o começo

*Eu aprendi qual é o valor de um sonho alcançar
Eu entendi que o caminho pedras terá
Eu vi em campo aberto se erguer construção
E foi com muitas pedras, e foi com muitas mãos*

*Eu vi o meu limite vir diante de mim
Eu enfrentei batalhas que eu não venci
Mas o troféu não é de quem não fracassou
Eu tive muitas quedas, mas não fiquei no chão*

*E ao olhar pra trás, tudo que passou
Venho agradecer quem comigo estava
Ergo minhas mãos pra reconhecer*

*E hoje eu sou quem eu sou
Pois Sua mão me acompanhava
Mas eu sei, não é o fim, é só o começo da jornada
Eu abro o meu coração pra minha nova história*

*Vejo vitórias e hoje eu olho pra trás
E a minha frente eu sei
Existem muito mais
Eu sei que minha jornada aqui só começou
Ao longo dessa estrada sozinho não estou*

Vocal Livre

RESUMO

Este trabalho tem como objetivo investigar as formas nominais usadas para referenciar o sujeito com referência indeterminada no português falado nas capitais nordestinas. A ancorou-se este estudo na Dialetologia (CARDOSO, 2010), na Sociolinguística (TARALLO, 2006; CEZARIO; VOTRE, 2008; LABOV, 2008) e nos estudos acerca da indeterminação da referência do sujeito (CUNHA; CINTRA, 1985; CEGALLA, 2008; ALMEIDA, 2009; ROCHA LIMA, 2011; BECHARA, 2004; MILANEZ, 1982; MENON, 2006 [1994]; DUARTE, 2007; AZEREDO (2008), SILVA (2004), PONTES, 1986; BARBOSA, 2016). O *corpus* que norteia esta análise foi coletado a partir de 72 entrevistas realizadas pelo Projeto ALiB – Atlas Linguístico do Brasil, com falantes nativos das localidades, distribuídos entre os fatores sexo (masculino e feminino), faixa etária I (de 18 a 30 anos) e faixa II (de 50 a 65 anos), e com dois níveis de escolaridade: sujeitos que cursaram, no máximo, o ensino fundamental e sujeitos que cursaram o ensino universitário. Os resultados indicam, no falar nordestino, o uso de diferentes estratégias para indeterminar a referência do sujeito, tais como: *O cara*, *O pessoal*, *O rapaz*, *Nego*, *Fulano*, dentre outras, e que o uso dessas FNs é motivado pelas variáveis sociais *localidade* e *sexo*.

Palavras-chaves: Indeterminação da referência do sujeito. Formas Nominais. Português falado no Nordeste. Dialetologia Pluridimensional. Sociolinguística.

RESUMEN

Este trabajo tiene como objetivo investigar las formas nominales utilizadas referirse al sujeto con referencia indeterminada en el portugués hablado en las capitales del Nordeste. Este estudio se basó en la Dialectología (CARDOSO, 2010), en la Sociolingüística (TARALLO, 2006; CEZARIO; VOTRE, 2008; LABOV, 2008) y en los estudios sobre la indeterminación de la referencia del sujeto (CUNHA; CINTRA, 1985; CEGALLA, 2008; ALMEIDA, 2009; ROCHA LIMA, 2011; BECHARA, 2004; MILANEZ, 1982; MENON, 2006 [1994]; DUARTE, 2007; AZEREDO (2008), SILVA (2004), PONTES, 1986; BARBOSA, 2016). El *corpus* de investigación de este análisis fue recolectado de 72 entrevistas realizadas por el Proyecto ALiB – Atlas Lingüístico de Brasil, con hablantes nativos de las localidades, distribuidos entre los factores sexo (masculino y femenino), grupo de edad I (de 18 a 30 años edad) e edad II (de 50 a 65 años), y con dos niveles de educación: sujetos que estudiaran hasta la escuela primaria y sujetos que estudió en la universidad. Los resultados indican, en habla de Nordeste, el uso de diferentes estrategias para indeterminar la referencia del sujeto, tales como: *O cara*, *O pessoal*, *O rapaz*, *Nego*, *Fulano*, entre otros, y que el uso de estas FNs son motivado por las variables sociales localidad y sexo.

Palabras clave: Indeterminación de la referencia del sujeto. Formas nominales. Portugués hablado en el Nordeste. Dialectología multidimensional. Sociolingüística.

LISTA DE ILUSTRAÇÕES

QUADROS

Quadro 1 - Principais diferenças entre indeterminação e indefinição de acordo com Milanez (1982).	38
Quadro 2 - Distribuição dos sujeitos da pesquisa, segundo o perfil do ALiB.	44
Quadro 3 - Variáveis analisadas	58

FIGURAS

Figura 1 - Rede de Pontos ALiB – Região Nordeste.....	48
Figura 2 - Planilha dos dados	55
Figura 3 - Divisão Dialetal proposta por Antenor Nascente	60

TABELAS

Tabela 1 - Distribuição total das ocorrências por grupo.....	67
Tabela 2 - Distribuição das ocorrências por localidade.....	71
Tabela 3 - Distribuição das FNs do Grupo A por localidade	73
Tabela 4 - Distribuição das FNs do Grupo B por localidade.....	75
Tabela 5 - Distribuição das FNs segundo sexo.....	76
Tabela 6 - Distribuição das FNs segundo o fator sexo	76
Tabela 7 - Distribuição das FNs segundo faixa etária	79
Tabela 8 - Distribuição das FNs segundo a faixa etária	80
Tabela 9 - Cruzamento entre sexo e faixa etária.....	83
Tabela 10 - Distribuição das FNs segundo a escolaridade	84
Tabela 11 - Distribuição das FNs segundo a escolaridade, em percentual.....	84
Tabela 12 - Cruzamento entre sexo e nível de escolaridade.....	87
Tabela 13 - Distribuição das FNs segundo Grau de Indeterminação	88
Tabela 14 - Distribuição das FNs segundo tempo verbal	90
Tabela 15 - Distribuição das FNs segundo Tempo verbal.....	91

GRÁFICO

Gráfico 1 - Distribuição dos recursos de indeterminação da referência do sujeito do Grupo A	68
Gráfico 2 - Distribuição dos recursos de indeterminação da referência do sujeito do Grupo B	69

SUMÁRIO

1. INTRODUÇÃO	14
2. O SUJEITO E A INDETERMINAÇÃO: ALGUMAS NOTAS.....	18
2.1 O sujeito e a indeterminação nas gramáticas normativas	18
2.2 O sujeito e a indeterminação nos manuais de linguística	26
2.3 Indeterminação no português brasileiro.....	31
2.4 As Formas Nominais (FNs)	35
2.4.1 Determinação, Indeterminação, Indefinição e Impessoalização	37
3. PROCEDIMENTOS METODOLÓGICOS	42
3.1 Projeto Atlas Linguístico do Brasil - ALiB	42
3.2 Perfil dos Informantes.....	43
3.3 Instrumentos da pesquisa	44
3.4 Região Nordeste: <i>Loci</i> da Pesquisa.....	45
3.4.1 São Luís (MA026).....	48
3.4.2 Teresina (PI034).....	49
3.4.3 Fortaleza (CE041)	49
3.4.4 Natal (RN053).....	50
3.4.5 João Pessoa (PB061)	51
3.4.6. Recife (PE070)	52
3.4.7. Maceió (AL077).....	52
3.4.8. Aracaju (SE079).....	53
3.4.9 Salvador (BA093)	54
3.5 O corpus e o tratamento quantitativo	54
3.6 Organização do envelope da variação.....	56
3.7 Escolhas dos fatores de análise	58
3.7.1 Fatores Sociais.....	58
a) Localidade	59
b) Sexo.....	61
c) Faixa Etária	61
d) Escolaridade	62
3.7.2 Fatores Linguísticos	63
a) Tempos verbais	63
b) Grau de indeterminação do sujeito	64

4. EXPOSIÇÃO E ANÁLISE DOS DADOS.....	66
4.2 Variáveis selecionadas pelo programa.....	71
4.2.1 Localidade.....	71
4.2.2 Sexo	75
4.3 Grupos de fatores não selecionados	79
4.3.1 Faixa etária	79
4.3.2 Escolaridade	83
4.3.4 Graus de indeterminação	88
4.3.5 Tempo verbal.....	90
5. CONSIDERAÇÕES FINAIS.....	93
REFERÊNCIAS	96

1. INTRODUÇÃO

O Português Brasileiro (doravante PB) possui diferentes estratégias para indeterminar a referência do sujeito na fala e na escrita, seja por não conhecer ou não querer determiná-lo. A ideia de não revelar a identidade de quem se fala está relacionado ao fato de impedir ou suavizar eventuais conflitos durante o processo de interação verbal. Assim, as estratégias de indeterminação do sujeito são recursos linguísticos que enveredam pelo âmbito semântico-pragmático da língua e envolvem situações contextuais extralinguísticas específicas, que por sua vez se valem das intenções comunicativas do falante.

As Gramáticas Tradicionais (doravante GTs), em sua maioria, trazem apenas as formas canônicas: verbo na terceira pessoa do plural sem sujeito lexical – da [\emptyset + 3ª pessoa plural] e verbo na terceira pessoa mais a partícula ‘se’ sem sujeito lexical – [\emptyset + verbo intransitivos + se] como recurso de indeterminação do sujeito. No entanto, estudos que tomam como base o português falado, incluem-se aí artigos científicos, gramáticas, dissertações e teses, nos apresentam um leque de possibilidades para indeterminar o sujeito. São, pois elas: verbo na terceira pessoa do singular sem sujeito lexical – [\emptyset + 3ª pessoa singular], verbo no infinitivo sem sujeito lexical – [\emptyset + infinitivo], verbo na primeira pessoa do plural, as formas pronominais *eu, você, nós, a gente e eles*, além das formas nominais.

As Formas Nominais (doravante FNs), objeto de estudo desta pesquisa, são formas com sujeito lexical usado semanticamente como referência genérica do sujeito. Para ilustrar, temos, a seguir, exemplos retirados de nosso *corpus*.

Ex. (1)

INQ. – ... a pessoa que tem os olhos voltados para direções diferentes? (Completar com um gesto dos dedos.)

INF. – Vesgo, é um termo bem comum vesgo, seria vesgo, se...

INQ. – Conhece também outro nome?

INF. – Ah! Tem, tem um nome, mais, pessoa que tem, é o termo correto?

INQ. – Não, eu digo mais assim um termo popular que você conhece?

INF. – Popular (inint) é, é vesgo, é tem também tem, instalação trocada. **FULANO** tem a instalação trocada, tem os olhos, estrábico, que seria o termo correto, é estrábico (rindo)¹. (MA23/7)²

¹ Os exemplos citados no trabalho seguem a transcrição grafemática, molde proposto pelo ALiB. As abreviaturas INQ. e INF. referem-se, respectivamente, a inquiridor e informante.

² A codificação dos sujeitos da pesquisa e das localidades encontra-se explicitada no capítulo Procedimentos Metodológicos.

Ex. (2)

INF. – ... A hente chama cisco

INQ. – E chama por outro nome?

INF. – Um arguero, o pessoal antigo chamaha muito. “Caiu um arguero no meu olho.”

INQ. – Mas, hoje, os mais novos não chamam...

INF. – Eh.

INQ. – Só chamam cisco?

INF. – Eh.

INQ. – Ou ainda chama argueiro?

INF. – Não, só **O PESSOAL** mais antigo, né, que fala. (RE070/4)

Como podemos ver nos exemplos (1) e (2), os informantes podem fazer uso de diferentes FNs, por exemplo, *Fulano, O pessoal*, para indeterminar a referência do sujeito. Convém dizer que o foco deste estudo é fazer um levantamento das formas de referências do sujeito, ou seja, para entender a *indeterminação do sujeito*, é necessário discutir a referência do sujeito, considerando aspectos semânticos, pragmáticos e sintáticos, principalmente no que concerne às formas nominais.

Sendo assim, seguindo o viés da Dialetologia Pluridimensional, estabelecemos como objetivo geral investigar as formas nominais de referência do sujeito, a fim de verificar a situação de uso na fala de nordestinos; e como objetivos específicos: (i) apresentar a variedade de formas nominais que os falantes das capitais nordestinas se valem para expressar a indeterminação do sujeito; e (ii) analisar os fatores sociais e linguísticos que influenciam ou mesmo condicionam as formas nominais de referência do sujeito na fala de nordestinos. Para desenvolver este estudo, escolhemos as nove capitais da Região Nordeste que integram a rede de pontos do ALiB, a saber: São Luís, Teresina, Fortaleza, Natal, João Pessoa, Recife, Maceió, Aracaju e Salvador.

Para tanto, levantamos algumas hipóteses que vão nos auxiliar no processo de análise dos dados. São elas:

1. As formas nominais mais preferidas entre os falantes das capitais nordestinas são *O pessoal / Os pessoal*, uma vez que há generalização do referente, ou seja, a ação verbal pode ser atribuída a qualquer pessoa.
 2. Há uma tendência de São Luís e Salvador terem mais diferentes tipos de *formas nominais de referência* (FNs) se comparado as outras localidades. Isso porque ambas as capitais estão situadas geograficamente mais próximas de
-

outras divisões dialetais, São Luís por se situar próximo às áreas do Falar Amazônico, e Salvador por estar próxima do Falar Fluminense.

3. Ainda que todas as variantes aqui analisadas se apresentem como inovadora, as mulheres terão maior frequência de uso das formas nominais que os homens, visto que elas são consideradas detentoras no processo de variação linguística.
4. A faixa etária mais avançada terá maior preferência pelas formas como *O pessoal/ A pessoa*, mesmo que as formas nominais sejam consideradas variantes inovadoras, se comparada às formas tradicionais apresentadas nas gramáticas tradicionais (a forma [\emptyset + 3ª pessoa singular], a forma [\emptyset + infinitivo]).
5. Mais anos de escolarização favorecem o uso de recursos para indeterminar a referência do sujeito.
6. As construções sintáticas no presente do indicativo favorecem o uso das formas nominais como recurso de indeterminação da referência do sujeito.
7. A indeterminação parcial com referência explícita no contexto tende a ser mais recorrente, visto que é possível encontrar no contexto qualquer elemento que auxilie na recuperação do referente.

Para a composição do *corpus* deste estudo, selecionamos os dados extraídos de 72 entrevistas realizadas pelo Projeto ALiB com falantes nativos das localidades, distribuídos segundo os fatores *sexo*, *idade*, *faixa etária* e *escolaridade*. (cf. Procedimentos Metodológicos).

Para a escolha dessas capitais, primeiramente levamos em consideração as Regiões Brasileira, para em seguida selecionar a Região Nordeste. Para tanto, foi pertinente considerar a divisão dialetal proposta por Antenor Nascentes, em seu livro *O linguajar carioca*, de 1922. Ao publicar essa obra, Nascentes levou em consideração as dificuldades que se tinha em ter uma divisão de áreas linguísticas no Brasil, ou seja, a falta de isoglossas³ precisas.

Assim, a proposta de Nascentes, que vem sendo ponto de partida para vários estudos de natureza dialetal que buscam examinar a atualidade da proposta e a pertinência dos limites nela estabelecidos, considerando a realidade linguístico-cultural do Brasil no momento atual, também é o ponto de partida deste trabalho. Nesse sentido,

³ Por isoglossas entende-se uma linha virtual que marca o limite, também virtual, de formas e expressões linguísticas (FERREIRA; CARDOSO, 1994, p.12-13).

acredita-se que seja oportuno verificar como um fenômeno morfossintático se comporta, atualmente, perante essa divisão dialetal do Brasil.

Para o nosso respaldo teórico, temos três momentos: no primeiro apresentamos o posicionamento de alguns gramáticos normativistas a respeito dos conceitos de sujeito e de indeterminação, assim, investigamos as obras de Cunha e Cintra (2001 [1985]), Cegalla (2008 [1977]), entre outros. No segundo momento, trouxemos autores que enfocam a indeterminação por uma perspectiva puramente linguística, considerando o contexto de fala, como Perini (1995) e Duarte (2007) e, por último, temos os resultados de algumas pesquisas que tomam como base dados orais de fala e que consideraram em sua análise os fatores linguísticos e sociais, como na pesquisa de Milanez (1982).

Visando explicitar os caminhos percorridos até a presente etapa da pesquisa, organizamos o trabalho da seguinte forma: (i) este tópico introdutório, com apontamentos sobre o objeto de estudo, os objetivos do trabalho e outra informações pertinentes para o desenvolvimento do nosso estudo; (ii) fundamentação teórica, com algumas considerações sobre o fenômeno em questão; (iii) procedimentos metodológicos, com a descrição das etapas seguidas; (iv) apresentação e análise dos dados; e (v) considerações acerca dos dados analisados.

2. O SUJEITO E A INDETERMINAÇÃO: ALGUMAS NOTAS

Neste capítulo iremos apresentar diferentes abordagens sobre a indeterminação do sujeito, começando pela normativa. Para tanto, tomaremos como base os apontamentos presentes nas seguintes gramáticas normativas: Cunha e Cintra (2001 [1985]), Bechara (2004), Cegalla (2008 [1977]), Almeida (2009) e Rocha Lima (2011). Além das normativas, analisamos a indeterminação da referência do sujeito em alguns manuais de linguística: Perini (1995; 2011), Pontes (1986), Silva (2004), Menon (2006 [1994]) e Duarte (2007) e Azeredo (2008). Por fim, verificamos os resultados de algumas pesquisas que usaram dados de falas de diferentes localidades: Milanez (1982), Barbosa (2016), Guimarães (2018).

2.1 O sujeito e a indeterminação nas gramáticas normativas

É fato consabido que as gramáticas normativas representam um papel importante na sistematização e na padronização das regras que organizam uma língua. O problema é que leva muito tempo para os gramáticos reconhecerem que a língua é como um elástico que se molda às situações comunicativas, e isso não seria diferente em se tratando do fenômeno da indeterminação do sujeito.

Assim, é necessário, como primeira tarefa, compreender a noção de sujeito, para em seguida analisar como a indeterminação é apresentada nas Gramáticas Tradicionais. Segundo Duarte (2007, p. 195), há uma incoerência ao se “misturar critérios sintáticos (estruturais) e semânticos” para se classificar os tipos de sujeitos. É possível encontrar nas GTs a seguinte classificação de sujeito: ‘simples’, ‘composto’, ‘oculto’, ‘indeterminado’, além da ‘oração sem sujeito’.

Segundo a referida autora, classificar o sujeito em “simples” e “composto” é desnecessário, tendo em vista que ambos podem apresentar “dois ou mais sintagmas coordenados”, da mesma forma é irrelevante a classificação de sujeito “oculto” (ou “subentendido”, “desinencial”, entre outros nomes) que só faz sentido se a ele se opuser o sujeito “expresso” (DUARTE, 2007, p. 195). Nesse sentido, notamos que as classificações aqui apresentadas são de natureza puramente sintática, voltando-se apenas para a estrutura da oração. Convém dizer que nosso objetivo neste tópico é mostrar as “inconsistências” em relação ao ponto de vista da GT a respeito da indeterminação da referência do sujeito e, a partir daí, assumir um posicionamento sobre o fenômeno em estudo.

Embora o conceito de sujeito gramatical seja constituído por uma mistura de critérios, é notório que prevalece o estudo no âmbito da sintaxe. Ao fazer uma revisão sobre o conceito de sujeito nas obras de Cunha e Cintra (2001 [1985]), Bechara (2004), Cegalla (2008 [1977]), Almeida (2009) e Rocha Lima (2011) constatamos que há uma problemática em relação à definição de sujeito.

Almeida (2009, p. 417) afirma que “sujeito de um verbo é a pessoa ou coisa sobre a qual se faz alguma declaração”. Na visão de Cunha e Cintra (2001 [1985], p. 119), “o sujeito é o termo sobre o qual se faz uma declaração”. Já Cegalla (2008 [1977], p. 324) salienta que “o sujeito é o ser do qual se diz alguma coisa.”. Notamos que o sujeito apresentado pelos autores configura-se como apenas um elemento gramatical na oração, sem considerar as questões de ordem semântica. Soma-se a isso o fato de essas definições de sujeito serem, na verdade, as conceituações de *tópico*⁴ e *comentário*⁵, que podem ou não coincidir com os elementos a que a gramática normativa denomina sujeito e predicado (DUARTE, 2007).

Rocha Lima (2011, p. 288), por sua vez, enfatiza que o sujeito é expresso por substantivo ou por elementos que possuem equivalência, podendo um único elemento exprimir a ideia de sujeito na oração. O autor segue dizendo que há casos em que é necessário algum elemento que possa restringir mais ainda a significação do substantivo que ocupa o lugar do sujeito na oração. Como é o caso do exemplo apresentado por ele:

Ex. (3)

“Branças pombas castíssimas voavam.”
(ROCHA LIMA, 2011, p.288, grifo nosso)

O referido autor salienta que o sujeito expresso na oração não é composto por apenas um substantivo, mas de um elemento que o acompanha, dando-lhe sentido fundamental. A respeito disso, dizemos que o substantivo *pomba* é o núcleo do sujeito, enquanto *brancas* é o elemento que acompanha o nome e restringe a característica do ser.

Ao pesquisar na obra de Bechara (2004), notamos que há em parte uma uniformidade com o descrito em Rocha Lima (2011), uma vez que a definição de sujeito apresentada por Bechara segue por critérios sintáticos, como podemos ver a seguir:

⁴ O *tópico* é definido como o sintagma nominal ou preposicional, externo à sentença, normalmente já ativado no contexto discursivo, sobre o qual se faz uma proposição por meio de um comentário. (MATEUS et alii, 2003).

⁵ O *comentário* é a informação nova dada pelo enunciador da sentença ao receptor. (PONTE, 1987)

[...] Chama-se sujeito à unidade ou sintagma nominal que estabelece uma relação predicativa com o núcleo verbal para constituir uma oração. É, na verdade, uma explicitação léxica do sujeito gramatical que o núcleo verbal da oração normalmente inclui como morfema número-pessoal [...] (2004, p.409, grifo nosso).

Para ilustrar o conceito apresentado em sua obra, o autor utiliza o seguinte exemplo:

Ex. (4):

Eu estudo no colégio e eu e dois irmãos brincamos no clube.
(BECHARA, 2004, p.409)

No exemplo dado pelo autor, temos como os núcleos verbais das duas orações **estudo** e **brincamos** que incluem os morfemas *o* (stud-o) e *mos* (brinca-mos), que indicam os sujeitos gramaticais, ou seja, 1ª pessoa do singular e 1ª pessoa do plural, respectivamente. Bechara (2004) salienta que o conceito de sujeito apresentado tem uma noção gramatical, não semântica. Reconhece que sujeito e paciente⁶ não são sinônimos, apenas uma concordância, além disso, a noção de sujeito pertence ao espaço sintático e “agente⁷” ao espaço semântico, ou seja, a relação que eles estabelecem entre si é meramente simbólica. Para tanto, o autor cita como exemplo as seguintes orações:

Ex. (5a):

“**Machado de Assis** escreveu extraordinários romances.”
(BECHARA, 2003, p.410, grifo nosso)

Ex. (5b):

“**Extraordinários romances** foram escritos por Machado de Assis”. (BECHARA, 2003, p.410, grifo nosso)

Em (5a), vemos que **Machado de Assis** ocupa os dois lugares o de agente e o de sujeito da oração; já na sentença (5b), **Machado de Assis** desempenha o papel de agente, no entanto não ocupa o lugar de sujeito, assim, temos como o sujeito paciente o sintagma nominal **Extraordinários romances**.

Embora Bechara (2004) tenha apresentado questão relacionada à agentividade, percebemos que os exemplos ora citados servem para corroborar a sua análise, de que os papéis agente e paciente não são de importância, mas sim a concordância entre o

⁶ *Paciente* se define como a entidade que sofre o efeito de uma ação ou evento, mudando de estado em consequência dela. (PERINI, 2006, p. 123)

⁷ *Agente* é o elemento que pratica uma ação. (PERINI, 2006, p. 123)

sujeito e o verbo. Logo, as questões de natureza semântica não são investigadas por este autor, tampouco pelos gramáticos já mencionados.

Como vimos, na perspectiva normativa, o tratamento da noção de sujeito não considera o contexto discursivo-pragmático, bastando, portanto, que exista uma forma linguística presente na estrutura que possa ocupar o lugar do sujeito. Assim, em estruturas como “Alguém tocou a campainha.”, o sujeito é considerado como determinado, mesmo que os interlocutores não tenham a referência de quem é esse *alguém*. Contrapondo-nos a essa ideia e apoiados em Franchi (2006), Duarte (2007), Perini (2006), dentre outros, pontuamos que o contexto discursivo e as propriedades semântico-pragmáticas (seu conteúdo e valor semântico) do sujeito constituem o eixo sobre o qual se assenta a noção de determinação/indeterminação da referência do sujeito.

Nessa perspectiva, é oportuno reforçar que a determinação/indeterminação é uma propriedade semântico-pragmática do sujeito, uma vez que todo sujeito apresenta um referente, estando ele determinado ou indeterminado a depender do contexto.

Em sua obra *Nova Gramática do Português Contemporâneo*, Cunha e Cintra (2001 [1985], p.125), após apresentar o conceito de sujeito, classifica-o em simples, composto, oculto (determinado) e indeterminado. O autor salienta ainda que a forma de representação de sujeito de 1ª ou de 2ª pessoa pode ocorrer por meio de pronomes (singular ou plural). Já os sujeitos da 3ª pessoa podem ter como núcleo um substantivo, os pronomes pessoais *ele(s)*, *ela(a)* no singular ou no plural, além de numeral e palavras ou expressão substantiva. Quanto ao sujeito indeterminado, o autor observa que para se indeterminar o sujeito é necessário que “algumas vezes o verbo não se refere a uma pessoa determinada, ou por se desconhecer quem executa a ação, ou por não haver interesse no seu conhecimento.” (CUNHA; CINTRA, 2001 [1985], p.125).

A respeito disso, Cunha e Cintra (2001 [1985]) enfatizam a existência de duas possibilidades para se indeterminar a referência do sujeito:

- a) Verbos na 3ª pessoa do plural.

Ex. (6):

“**Reputavam-no** o maior comilão da cidade.”
(CUNHA; CINTRA, 2001 [1985], p. 94).

b) Verbo na 3ª pessoa do singular, com o pronome *se*:

Ex. (7):

“Ainda **se** vive num mundo de certezas.”
(CUNHA; CINTRA, 2001 [1985], p. 94).

Notamos que os exemplos dados pelo gramático referem-se a sujeitos implícitos, ou seja, não expresso materialmente, além de serem verbos intransitivos e transitivos diretos. Vemos que o conceito apresentado permite uma dupla interpretação, uma vez que o verbo não fornece meios que nos ajude a identificar quem “executa a ação”, podendo ser: autor/falante ou leitor/ouvinte.

Na *Novíssima Gramática da Língua Portuguesa*, Cegalla (2008) enfatiza que o sujeito pode ser expresso na oração por substantivo, pronome, palavra ou expressão. Ressalta ainda que o núcleo do sujeito, ou seja, a palavra-base pode ser ocupada por um único substantivo ou pronome. No entanto, em torno do núcleo podem aparecer palavras secundárias (artigos, adjetivos, locuções adjetivas, etc.). O autor segue apresentando oito tipos de sujeito:

1. *simples* (quando tem um só núcleo);
2. *composto* (quando tem mais de um núcleo);
3. *expresso* (quando está explícito);
4. *oculto* ou *elíptico* (quando está implícito; quando não está expresso, mas se deduz através do contexto);
5. *agente* (se faz a ação expressa pelo verbo da voz ativa);
6. *paciente* (quando sofre ou recebe os efeitos da ação expressa pelo verbo passivo);
7. *agente e paciente* (quando o sujeito faz a ação expressa por um verbo reflexivo e ele mesmo sofre os efeitos dessa ação);
8. *indeterminado* (quando não se indica o agente da ação verbal).

Como já mencionado, Cegalla (2008) mistura a noção sintática de sujeito com o conceito de agente e paciente, no entanto, para determinar e quantificar o núcleo e a expressão lexical ou não do sujeito, o autor utiliza critérios morfossintáticos. O uso dos critérios semânticos (agente e paciente) objetiva estabelecer uma regularidade entre estrutura sintática e semântica, uma vez que podemos ter estrutura ativa ou passiva na

oração, mesmo que estruturas de voz passiva não seja algo exclusivo do sujeito paciente.

A respeito do conceito de *sujeito indeterminado*, o referido autor introduz o assunto conceituando-o, para, em seguida, apresentar as três possibilidades que podem ser usadas.

- a) Verbo na 3ª pessoa do plural, sem referência a qualquer agente já expresso nas orações anteriores.

Ex. (8):

“Bateram palmas no portãozinho da frente.”
(CEGALLA, 2008, p.325).

- b) Verbo ativo na 3ª pessoa do singular, acompanhado do pronome se.

Ex. (9):

“Aqui se vive bem.”
(CEGALLA, 2008, p. 325).

- c) Verbo no infinitivo impessoal.

Ex. (10):

“Era penoso carregar aqueles fardos enormes.”
(CEGALLA, 2008, p. 325).

Podemos perceber que as possibilidades apresentadas por Cegalla (2008) não diferem muito das apresentadas por Cunha e Cintra (1985), no entanto, Cegalla (2008) acrescenta mais uma forma para se indeterminar o sujeito na oração, a de verbos no infinitivo pessoal, ou seja, quando o verbo expressa uma ideia vaga sem indicar o sujeito que executou a ação.

Outro ponto importante que Cegalla (2008) destaca é que não devemos confundir sujeito indeterminado com sujeito oculto, uma vez que para se deduzir o sujeito que não está expresso é necessário do contexto a qual está inserido. No entanto, para analisar sintaticamente o sujeito indeterminado é indispensável a presença no contexto para que seja possível recuperar a referência de quem está falando. Logo podemos notar inconsistências sobre esse ponto destacado pelo autor. Cegalla (2008) menciona também que sujeito formado por pronome indefinido não é indeterminado,

mas expresso, como em “Alguém me ensinará o caminho.” ou “Ninguém lhe telefonou.”

Na visão apresentada na obra *Gramática Metódica da Língua Portuguesa*, de Almeida (2009), notamos que o autor classifica o sujeito em três tipos: simples, composto e indeterminado. A respeito do conceito de sujeito indeterminado, podemos ver que não se distancia muito das definições apresentadas por Cunha e Cintra (2001 [1985]) e Cegalla (2008). Para Almeida, sujeito indeterminado é “quando de impossível identificação” e cita duas alternativas para se identificar esse tipo de sujeito (ALMEIDA, 2009, p. 414).

a) Ativo, acidentalmente impessoalizado na 3ª do plural.

Ex. (11):

“Dizem que ele vem.”

(ALMEIDA, 2009, p.414).

b) Acidentalmente impessoalizado na passiva.

Ex. (12):

“Precisa-se de um datilógrafo.”

(ALMEIDA, 2009, p.414).

Vemos na definição apresentada por Almeida (2009) um extremismo ao usar o termo “impossível de identificação”, uma vez que podemos sim recuperar a referência do sujeito através do contexto. Em outra seção, o autor terce comentários sobre as diversas funções do pronome “se”, dentre elas salienta que tal pronome apresenta um caráter de impessoalidade quando acompanha verbos intransitivos ou verbos transitivos indiretos, no entanto é necessário que o verbo esteja no singular para que isso aconteça, uma vez que, segundo o autor, é “dizer tolice em português” (ALMEIDA, 2009, p. 2018).

Na *Gramática Normativa da Língua Portuguesa*, de Rocha Lima (2011), notamos que após introduzir o conceito de sujeito, diferente de Cunha e Cintra (2001 [1985]), Cegalla (2008) e de Almeida (2009), o referido autor não estabelece uma classificação para os tipos de sujeito, apenas o divide em determinado ou indeterminado. Salienta que é determinado o sujeito que for possível se identificar na oração, sendo explícito ou implícito; já o indeterminado, se não pudermos ou não

quisermos especificá-lo. Orienta que para indeterminar o sujeito, vale-se a língua de um dos recursos:

- a) Empregar o verbo na 3ª pessoa do plural, sem referência anterior ao pronome eles ou elas, e a substantivo no plural;

Ex. (13):

“**Falam** mal daquela moça.”
(ROCHA LIMA, 2011, p.289)

Ex. (14):

“**Mataram** um guarda.”
(ROCHA LIMA, 2011, p.289)

- b) Usá-lo na 3ª pessoa do singular acompanhado da partícula *se*, desde que o verbo seja intransitivo, ou traga complemento preposicional.

Ex. (15):

“**Vive-se** bem aqui.”
(ROCHA LIMA, 2011, p.289)

Ex. (16):

“**Precisa-se** de professores.”
(ROCHA LIMA, 2011, p.289)

Como podemos ver o conceito de *sujeito indeterminado* apresentado por Rocha Lima (2011) não se difere muito dos outros já mencionados neste trabalho, contudo percebemos que ele não se apega os tipos – *simples, composto e indeterminado* – mas sim à noção de determinação e indeterminação. Quantos aos recursos usados para indeterminar a referência do sujeito, o autor segue com as mesmas formas já difundidas pelos gramáticos tradicionais, assim como os exemplos sem contexto, apenas orações isoladas.

Bechara (2004), por sua vez, apresenta em sua obra *Moderna gramática portuguesa* o conceito de *sujeito indeterminado* e salienta que “muitas vezes não sabemos ou não queremos determinar com precisão o sujeito da oração” (BECHARA, 2004, p. 409), no entanto, é sabido que a ação foi praticada por um ser que pode ou não ser apresentado no discurso. Ele também aponta as duas formas possíveis, na língua portuguesa, para indeterminar o sujeito:

- a) Emprego de verbos na 3ª pessoa (singular ou plural) sem referência a pessoas determinadas, sendo mais usual a forma que se apresenta no plural.

Ex. (17):

“Estão batendo.”

(BECHARA, 2004, p.31)

- b) Emprego do pronome *se* junto a verbo de modo que a oração passe a equivaler a outra que tem por sujeito *alguém, a gente*:

Ex. (18):

“Vive-se bem aqui. > A gente vive bem aqui.”

(BECHARA, 2004, p.31)

Ex. (19):

“Precisa-se de empregados. > Alguém precisa de empregados.”

(BECHARA, 2004, p.3, grifo nosso)

Notamos que o exemplo usado por Bechara (2004) traz o verbo *estar* na terceira pessoa do plural, corroborando com o que foi pontuado anteriormente. Vemos também que não conseguimos ter acesso à referência da pessoa determinada e isso ocorre porque não temos o contexto para recuperá-la.

Nos exemplos apresentados pelo autor, é possível ver que não há recurso para se recuperar a referência do sujeito. Ocorre a reestruturação das orações, com a substituição do primeiro “se” pelo pronome “a gente”, que também pode ser usado como elemento para se indeterminar o sujeito.

Conforme as definições e exemplos aqui apresentados pelos estudos tradicionais no PB, notamos que há uma mistura de critérios semântico-pragmáticos e sintáticos para definir o conceito de sujeito, o que dificulta o entendimento do termo empregado. Vemos que quase sempre o critério sintático é priorizado nas análises. Com base nessas informações, apresentaremos na seção, a seguir, algumas notas sobre como o conceito de sujeito e indeterminação são explorados por autores que embalam suas análises numa visão mais reflexiva da língua.

2.2 O sujeito e a indeterminação nos manuais de linguística

Estudos que embalam as discussões encontradas em muitos manuais de linguísticas partem de críticas feitas à gramática normativa e se fortalecem por meio de

estudos realizados com dados de fala e escrita. São esses manuais que nos possibilitam uma visão mais abrangente dos fenômenos linguísticos, como é o caso da indeterminação da referência do sujeito. De modo geral, as GTs conceituam sujeito como algo ou alguém de quem ou do que se fala. Como vimos, esse conceito apresenta-se de forma equivocada, uma vez que há mistura de critérios semântico e sintático.

Para Perini (1995), a noção de sujeito vai muito além do ser que praticou ou recebeu a ação, visto que o ponto de partida para sua investigação surge com base na organização da estrutura oracional, ou seja, torna-se precioso verificar cada um dos constituintes que se divide a oração e como eles estabelecem relação entre si. Há casos em que terão comportamento semelhante, mas poderão desempenhar funções diferentes.

Em se tratando do conceito de sujeito, Perini (1995, p.77) o define como “o termo da oração que está em relação de concordância com núcleo do predicado”. Convém dizer que o conceito apresentado pelo autor é uma definição formal e não leva em consideração o papel semântico ou discursivo do termo em questão, ou seja, o objetivo não é estabelecer o agente da ação, tampouco saber sobre alguma entidade a qual se declara alguma coisa, mas sim identificar um dos constituintes que compõem a oração e analisar a relação formal com outros elementos.

Analisar a relação que o sujeito estabelece com outros elementos está intimamente ligado com os traços que o constituinte tem na oração, ou seja, a relação de concordância verbal [+CV], como é o caso do exemplo citado a seguir:

Ex. (20):

“Meus sobrinhos comeram a melancia.”
(PERINI, 1995, p.76)

Ex. (21):

“Meus sobrinhos comeram a melancia.”
(PERINI, 1995, p.76)

No exemplo citado pelo autor vemos que o sintagma nominal *meus sobrinhos* é marcado pelo traço [+CV], entretanto os outros elementos que compõe e não estabelecem relação com o núcleo do predicado são marcados pelo traço [-CV]. Logo, analisar sintaticamente uma oração está relacionado aos traços atribuídos a determinado constituinte e como eles se apresentam. É importante ressaltar que focar apenas nessa perspectiva de análise pode complicar a conceituação de sujeito, uma vez que não podemos pura e simplesmente dizer que sujeito se trata do termo que estabelece relação

de concordância com o verbo. No entanto, essa visão apresentada por Perini (1995) nos possibilita analisar as condições de proposições de sujeito, além de verificar os pronomes retos de modo mais simplificado.

Em se tratando do tema indeterminação do sujeito, Perini (2010) em sua outra obra intitulada *Gramática do português brasileiro*, elenca os seguintes recursos sintáticos e lexicais para se indeterminar a referência do sujeito:

- a) sintagma nominal sem determinante.

Ex. (22):

“Criança suja muito o chão.”
(PERINI, 2010, p.84)

- b) verbo na terceira pessoa do plural desacompanhado de sujeito.

Ex. (23):

“Quebraram a janela.”
(PERINI, 2010, p.84)

- c) verbo sem sujeito na terceira pessoa do singular.

Ex. (24):

“Nessa fazenda planta café e milho.”
(PERINI, 2010, p. 85)

- d) verbo no infinitivo sem sujeito

Ex. (25):

“Nadar é bom para a saúde.”
(PERINI, 2010, p.85)

- e) uso de determinados itens lexicais.

Ex. (26):

“Ou você pertence àquele grupo de críticos ou te ignoram completamente.”
(PERINI, 2010, p. 85)

Ex. (27):

“O sujeito toma droga e ameaça quebrar tudo.”
(PERINI, 2010, p. 85)

Vemos que Perini (2010) nos apresenta possibilidades para indeterminarmos a referência do sujeito, deixando evidente que a visão demonstrada pelas GTs é limitada no sentido de nos apresentar poucas possibilidades se comparadas aos manuais. Em nota de rodapé, ele ainda admite ser rara, no PB, construções com o pronome “se”, corroborando com o pensamento de que nem tudo que encontramos nas gramáticas normativas devem ser encaradas como parâmetro a ser seguido.

Outra autora que expõe sua opinião sobre o assunto aqui debatido é Ponte (1986) que assinala que para se definir sujeito, faz-se necessário estabelecer uma distinção entre sujeito e agente, assim como fez Perini (1995). Nesse caso a autora pontua que “[...] o sujeito é afetado e não o agente [...] o sujeito sofre a ação” (PONTES, 1986, p.21). Reforça a ideia de que sujeito é um conceito sintático, ao passo que agente e paciente são conceitos semânticos, podendo estes concordar ou não na oração. Salieta ainda que o traço de agente não deve ser levando em consideração ao se estabelecer um conceito de sujeito. A respeito disso, a autora apresenta os seguintes exemplos:

Ex. (28):

“O carro furou o pneu.”
(PONTES, 1986, p.17).

Ex. (29):

“Eu opereí o estômago.”
(PONTES, 1986, p.20).

Nos exemplos acima apresentados, é possível encontrar traços de um sujeito básico, ou seja, agente, animado, tópico, pré-verbal. Vemos também que a autora enfatiza que é necessária uma atenção maior com sentença de ordem sujeito-verbo e objeto (SVO), pois estas permitem que interpretações equivocadas reflitam a presença do sujeito-agente, quando na verdade não tem. Notamos que Pontes (1986) corrobora com o pensamento de Cunha e Cintra (2001 [1985]) quanto à atitude do sujeito e a referência que ele estabelece com os verbos de ação.

Pontes (1986) observou que o problema de estabelecer um conceito para sujeito não distingue-se muito dos outros problemas conceituais da linguística. Buscou analisar diferentes vertentes teóricas e propôs uma visão que seguisse aos moldes do sujeito típico, haja vista que seria mais fácil de entender e fugir daquilo defendido pelas GTs.

Sob essa perspectiva, Azeredo (2008), por sua vez, apresenta de forma sintética sua visão sobre sujeito indeterminado:

Orações de “sujeito indeterminado” são empregadas por motivos cognitivos ou discursivos variados, e a língua oferece a seus usuários diferentes meios para indeterminar, dissimular ou mesmo ocultar a identidade do ser humano a quem o sujeito da oração se refere. A razão cognitiva óbvia é o desconhecimento da identidade do ser de que se fala. As razões discursivas, por sua vez, são variadas: a conveniência ou oportunidade da omissão da identidade do sujeito é uma delas, o registro de linguagem empregado ou o gênero de texto produzido é outra. (AZEREDO, 2008, p. 225).

Assim, o autor sinaliza no exemplo (30) formas mais canônicas de indeterminação e reconhece que, em situações mais espontâneas, o falante pode fazer uso de verbos na 3ª pessoa do singular, mesmo sem utilizar a partícula “se”, além de usar sintagmas nominais de natureza mais genérica em que não seja possível determinar o sujeito do enunciado. Para os sintagmas nominais, Azeredo (2008) classifica como sujeito simples.

Ex. (30):

“Estão anunciando na televisão um tratamento infalível para a calvície.”
(AZEREDO, 2008, p. 226).

Ainda seguindo esse viés, mas em uma perspectiva Sociolinguística, com base em dados de falas do PB, encontramos, por exemplo, o estudo de Menon (2006 [1994]), intitulado *Analyse sociolinguistique de l'indétermination du sujets dans lê portugais parlé au Brésil, à partir des données du NURC/SP*, que afirma:

Para mim, indeterminação do sujeito concerne os casos em que não se pode ou não se quer nomear o sujeito, na acepção de ‘referente extralinguístico’. No entanto, o referente é conhecido pelo locutor (e em certos casos, também do interlocutor, o que torna possível a compreensão mútua) e se ele quisesse ou se isso lhe fosse conveniente ou interessante, ele poderia nomeá-lo ou descrevê-lo. Nesse sentido, o referente pode ser recuperado pelo locutor a qualquer hora. Trata-se, antes de tudo, de uma maneira de escamotear o sujeito extralinguístico por meio de uma forma de expressão linguística, em função da situação de comunicação. (MENON, 2006 [1994], p. 129)

Convém salientar que estudos que põem em xeque a visão normativista de indeterminação da referência do sujeito e prescrição de apenas três formas para indeterminá-lo – (i) o uso do verbo na 3ª pessoa do plural, sem fazer referência ao sujeito expresso anteriormente no contexto; (ii) verbo (ativo) na 3ª pessoa do singular, acompanhado da partícula ‘se’; e (iii) com o uso do verbo no infinitivo impessoal – são relativamente recentes, já que são os estudos com base na língua falada, e não na escrita

literária, que enfocaram o tema como o fenômeno variável, registrado em diferentes organizações discursivas.

A esse respeito, Silva (2004, p. 80) afirma que as formas expressas de 1ª pessoa do plural “podem abarcar uma variedade de referenciais, específicos e genéricos, elaborados nas negociações entre interlocutores”. Milanez (1982), por sua vez, salienta que a forma como a pessoa do discurso é referenciada interfere na propriedade de determinação da referência do sujeito, e assinala ainda que

[...] estamos, portanto, diante de um processo da língua que permite ao falante passar do universo das três pessoas especificadas e identificáveis (que consideraremos como o nível da determinação) a um nível de generalização, que transcende o anterior por implicar numa referência de tal forma abrangente que pode envolver qualquer pessoa. É o fenômeno da indeterminação. (MILANEZ, 1982, p. 26).

Assim, discussões que envolvam a indeterminação da referência do sujeito são pertinentes, uma vez que qualquer uma das pessoas do discurso pode ocupar o lugar de sujeito na oração, bem como outros elementos, por exemplo, as formas nominais. Vale destacar que “os sujeitos de referência ‘indeterminada’ são hoje preferencialmente ‘expressos’ na fala, seja pelo pronome ‘eles’ seja por ‘a gente’ e principalmente por ‘você’.” (DUARTE, 2007, p. 196), além de formas nominais como *o povo*, *as pessoas*, *a galera*, *o pessoal*, por exemplo.

Sobre isso, apresentaremos, no tópico a seguir, algumas das pesquisas que investigam em diferentes comunidades linguísticas o fenômeno na indeterminação da referência do sujeito. Os trabalhos que citaremos analisam as formas tradicionais, aquelas apresentadas nas GTs, e as consideradas inovadoras, por exemplo, as formas nominais.

2.3 Indeterminação no português brasileiro

No Brasil, a indeterminação da referência do sujeito tem sido estudada à luz de diferentes vertentes linguísticas, principalmente sob a lente da Sociolinguística. Com o objetivo de mostrar um panorama geral das investigações a respeito da indeterminação do sujeito no PB, de modo que possamos visualizar como esse fenômeno tem se comportado em algumas localidades do país, apresentamos, a seguir, alguns estudos sobre o fenômeno em questão, no entanto priorizando as formas nominais, objeto de estudo deste trabalho.

Milanez (1982), em sua dissertação de mestrado, investiga o fenômeno da indeterminação partindo dos dados de gravações do *Projeto de Estudo da Norma Urbana Linguística Culta* (NURC), de São Paulo. A autora justifica seu trabalho na observação de que se ensina nas escolas sobre a indeterminação é algo muito restrito e muito distante da atuação oral do falante. Embora não seja da área da Sociolinguística, Milanez enfatiza que é importante compreender o fenômeno e reconhecer que a língua muda conforme as situações comunicativas a qual estamos inseridos. Sua linha de análise segue pela Gramática Gerativo Transformacional, valendo-se também da Semântica e da Pragmática, visto que o processo de indeterminação vai além dos critérios sintáticos, pois é necessário a análise do contexto de uso para compreendê-lo. Além disso, ela embasa sua discussão teórica com outros dois conceitos além do de *indeterminação*, os de *impessoalização* e a *indefinição*.

Na exposição dos dados, Milanez (1982) os divide em dois grandes grupos: formas sem sujeito lexical e formas com sujeito lexical. No primeiro grupo podemos encontrar as seguintes formas: o uso do verbo na terceira pessoa do plural; verbo na terceira pessoa do singular (com e sem a presença da partícula 'se'), já no segundo grupo temos: *a gente*; *você*; *eles*; *eu*; *nós*; *o/um indivíduo*; *o/um sujeito*; *a/uma pessoa*; e *o/um cara*.

De forma geral, foi computado pela autora um total de 1058 ocorrências, sendo 38% para as formas sem sujeito lexicalizado e 62% com formas lexicais, com uma parte significativa para as formas pronominalizadas. Algumas formas lexicais se mostraram recorrentes no *corpus*, tais como: *a gente* e *você*. O uso dessas formas se justifica pelo contexto a qual o falante está inserido, mesmo sendo de forma mais generalizada e informal.

Quanto às formas nominais apresentadas por Milanez (1982), a autora não se aprofunda muito sobre o assunto, deixando evidente que as formas *a gente*, *você* e *eles* foram mais representativas para a amostra. Enfatiza que o uso das formas *o/um indivíduo*; *o/um sujeito*; *a/uma pessoa*; e *o/um cara* computaram apenas 9% do total dos dados. Conclui dizendo que há um distanciamento significativo entre o que é ensinado nas escolas e o que é característico da fala e que o fenômeno da indeterminação se dá no nível semântico-pragmático da língua, explicável pelo contexto a qual está inserido.

Barbosa (2016) estuda algumas possibilidades de indeterminação da referência do sujeito e, para tanto, tomou como base dados de fala de 48 informantes de São Luís, Capital do Maranhão, estratificado de acordo com o sexo, faixa etária e escolaridade dos

falantes. O objetivo do seu estudo foi investigar as estratégias de indeterminação do sujeito utilizadas no português ludovicense, a fim de verificar o comportamento linguístico dessa comunidade de fala específica, além verificar se há contextos sociais que influenciam no uso dos recursos indeterminadores. A autora salienta que além das duas possibilidades apresentada pela gramática tradicional, é possível encontrar no PB construções com referência indeterminada com: *eu, tu, você, a gente, nós, eles, formas nominais, vocês, se, voz passiva sem agente, voz passiva sintética, se e verbo na terceira pessoa do plural*. Convém dizer que Barbosa (2016) não registrou em sua dissertação todas as possibilidades de recursos para indeterminar a referência do sujeito, apresentando um quantitativo de 10 formas.

Segundo a autora, foram analisados no total 3.804 dados de *sujeito indeterminado*. A respeito das variantes mais recorrentes, ela destaca as seguintes: *você* (1.127 ocorrências), *Verbo na terceira pessoa do singular* (872 ocorrências) e as *FNs* 560 (ocorrências). Foi constatado nas análises apresentada pela autora grande influência dos fatores sociais sob as estratégias de indeterminação do sujeito, dando destaque à escolaridade e o sexo dos informantes para variáveis que mais se mostraram recorrente.

Em relação às *FNs*, Barbosa (2016) constatou que há diferença significativa entre os falantes que possuem ensino fundamental e os falantes com ensino médio e universitário. Destaca que os dois primeiros níveis de escolaridade apresentaram resultados iguais para o desfavorecimento do uso das *FNs*, já os falantes com ensino universitário, demonstraram uma leve tendência ao uso dessas formas.

Na sua dissertação, a autora dedica uma seção às *FNs*, observamos o registro de 15 formas (*A pessoa, As pessoas, O pessoal, O cara, Nego, O povo, Neguinho, Pessoal, Neguin, Pessoa, A galera, Cara, Os cara, Negada e O camarada*.) para se indeterminar a referência do sujeito. Sobre o cômputo geral das *FNs*, foram registradas 560 ocorrências. Quanto à distribuição das *FNs* por estratégia, notamos que *A pessoa, As pessoas e O pessoal* foram mais recorrentes se comparada as outras. Em síntese, a autora reforça dizendo que a língua não acontece de modo aleatório, e que existe uma sistematização no tocante à variação linguística. Reforça ainda dizendo que seu trabalho representa de modo mais palpável a realidade linguística dos falantes de São Luís, visando, assim contribuir para o entendimento do português falando no Maranhão.

Guimarães (2018) se propôs a investigar as formas de indeterminação do sujeito no Estado da Bahia, mais especificamente em sete cidades: Barreiras, Barra, Vitória da Conquista, Ilhéus, Irecê, Alagoinhas e Salvador. Para seu estudo, tomou

como base os dados de fala de 28 informantes do *Projeto Atlas Linguístico do Brasil*, distribuídos entre sexo, faixa etária e escolaridade. Para análise, a autora considerou dividir as variantes em três grupos: grupo 1 - formas gramaticais; grupo 2 - formas nominais; grupo 3 - formas pronominais, seguindo os critérios apresentados pela teoria laboviana, de analisar as variáveis linguísticas e sociais.

O cômputo geral dos dados nos mostra o total de 811 ocorrências, sendo com maior registro para as formas pertencentes ao grupo 1, com 403 realizações. Nesse grupo, a autora relata que o uso do verbo na terceira pessoa do singular, desacompanhado da partícula ‘se’ se mostrou mais relevante. Quanto ao grupo 2, composto pelas formas nominais, foi registrado 256 ocorrências; já o grupo 3 das formas pronominais, teve o total de 143 dados.

Sobre o grupo 2, Guimarães (2018) subdivide em formas nominais A, B e C, cujos dados obedecem a uma mesma estrutura linguística. No subgrupo A são compostas FNs com as seguintes estruturas: artigo acompanhado do substantivo ou pronome indefinido “*As pessoas falam trovoadas/ os outros usando*” (p.71); o subgrupo B são estruturas formadas por pronome indefinido seguido de substantivo ou de pronome relativo/interrogativo “*Todo mundo aqui chama assim*” (p.17) e o subgrupo C, com um único elemento “*Fulano tá de boi.*” (p.17).

A respeito dos resultados encontrados, a referida autora salienta que as variantes nominais do grupo A apresentaram maior número de ocorrência, computando 148 dados, e sinaliza que foram encontradas estruturas como *As pessoas, Os outros, O povo, O pessoal, A maioria, Os cara(s) e A turma*, seguido do grupo B com 106 realizações e o grupo C, com 16 ocorrências.

Em síntese, Guimarães (2018) conclui seu estudo dizendo que os informantes pertencentes à faixa etária I desfavoreceram mais o uso das formas de indeterminação do sujeito que os da faixa etária II. Quanto aos tipos de variantes nominais, a faixa etária também se mostrou relevante, bem como a variável “cidade/mesorregião” que representou um fator condicionante à escolha de formada nominal presente no grupo A (artigo + nome ou pronome indefinido). Conclui ainda dizendo que trabalhos futuros podem confirmar ou não que foi apresentado em sua dissertação.

Convém ressaltar que os trabalhos ora citados investigaram as mais variadas formas de *indeterminação do sujeito* (gramaticais, nominais e pronominais), diferente do foco do nosso estudo, que tem como objetivo investigar apenas as formas nominais

nas capitais da Região do Nordeste. A respeito das formas nominais, apresentamos no tópico a seguir como o fenômeno da indeterminação é visto sob a nossa perspectiva.

2.4 As Formas Nominais (FNs)

Diante dos conceitos expostos nesta pesquisa, vimos que não é uma tarefa fácil definir a indeterminação do sujeito, uma vez que, segundo as palavras da autora Milanez (1982, p. 23), “a indeterminação é um fenômeno complexo que dá margens a estudos extensivos em diferentes níveis”

Como sinalizado na Introdução deste trabalho, nosso objetivo é estudar as formas nominais (FNs), ou seja, as formas compostas por [artigo definido + um substantivo] ou por um único elemento no sentido generalizante, usados para indeterminar a referências do sujeito. Segundo Menon (2006 [1994]), inicialmente essas formas eram usadas como locuções nominais comuns, porém, com o passar do tempo, elas foram se distanciando do seu significado original e assumindo unidades mais cristalizadas com significados mais genéricos.

Para ilustrar, trouxemos do nosso *corpus* alguns exemplos que sinalizam o que foi apresentado.

Ex. (31)

INQ.– E aquele clarão que surge no céu em dias de chuva? Você já falou de raio, mas não é o raio, ele é...aliás ele é...acho que é antes do raio. É um clarão que aparece, assim, de repente.

INF.– **O POVO** chama relâmpago. (MA023/8)

Ex. (32)

INQ. – Um acontecimento, uma coisa assim que você se lembre, ou...

INF. – Olhe, uma coisa que tá acontecendo que é importante pra mim e pro meu irmão...

INQ. – Certo.

INF. –... é uma viagem que ele que fazê muito, a trabalho. Isso aí... eu tô muito feliz por casa disso.

INQ. – É, né?

INF. – Com certeza.

INQ. – E... como é que foi essa viagem? Como (inint)?

INF. – Não, ele vai agora em dezembro. Ele vai mais um rapaz... pra Fernando de Noronha... trabalha lá.

INQ. – E é um trabalho interessante...

INF. – É... ele que muito.

INQ. – Vai trabalhar de quê?

INF. – Eu num sei dizê não. Eh... **O RAPAZ** vai levá ele, que... esse rapaz tem pousada lá, sabe? (RE070/4)

Ex. (33)

INQ. – Seu Julho, você acha que tem gente que fala diferente aqui em Natal?

INF. – Tem criatura, de ôtas... ôtas nacionalidades.

INQ. – Gente que tenha nascido no Rio Grande Norte, mas que o senhor acha que fala diferente.

INF. – Fala mais vexado. **PESSOAS** que falam vexado. Que a gente não entende. Tem pessoas que fala numa calma, que a gente precisa tê aquela paciência pra escutá direito. (RN053/3)

Ex. (34)

INF. – Cabriola. (inint) chama cabriola

INQ. – Uhn.

INF. – Sai rodano assim.

INQ. – É.

INF. – Chega e cai sentado.

INQ. – Isso.

INF. – **NEGO** chama cabriola. (BA093/3)

É possível ver nos exemplos apresentado a existência do sujeito expresso por diferentes FNs, no entanto não temos o referente participante do discurso, logo a referência desse sujeito torna-se indeterminada, visto que o contexto não permite que se identifique com precisão “quem chama relâmpago”, “quem é o rapaz que vai lá”, “quem é/são as pessoas que fala vexado” ou “quem chama cabriola”, podendo referir-se a qualquer pessoa do discurso. Notamos que o contexto a qual as FNs aqui apresentada estão possuem um conteúdo generalizado, ou seja, engloba as três pessoas do discurso, sendo necessário que o contexto ajude a recuperar parte dessa referência perdida.

Convém dizer que *a gente* não pode ser considerada uma forma nominal, embora possua a mesma composição. Isso se dá pelo fato de que a forma nominal *gente* ter passado pelo processo de gramaticalização, perdendo traços de gênero [+ feminino], e ganhando traços de [+ pessoa], tornando-se um pronome. Assim, com base no que pontua Menon (2006 [1994]), adotaremos o mesmo posicionamento.

Como salienta Milanez, para analisar a indeterminação, e, conseqüentemente, defini-la é necessário que se leve em consideração a noção dos seguintes elementos: pessoa, a de generalização e a de referência. Cavalcante (1998) enfatiza que das três a que mais se destaca é a noção de referência, pois é nela que a noção de pessoa e generalização se assenta. Assim, tendo em visto o exposto por Milanez (1982) e

Cavalcante (1998), concordamos que com base nesses apontamentos podemos delimitar um conceito para indeterminação.

2.4.1 Determinação, Indeterminação, Indefinição e Impessoalização

Em busca de melhor entendimento a respeito do fenômeno da indeterminação do sujeito, as autoras Milanez (1982) e Cavalcante (1998) apresentaram em seus estudos elementos que auxiliam na delimitação dos conceitos de indeterminação entre a determinação, indefinição e a impessoalização, uma vez que esse fenômeno apresenta características complexa no que tange a sua análise, sendo possível ser estudado em diferentes níveis. As autoras expõem por meios de suas pesquisas posicionamentos próprios que evidenciam elementos que são caracterizadores da indeterminação e que são de suma importância para delimitar o objeto em estudo.

Em se tratando da diferença entre determinação e indeterminação, é possível ver que as GTs não se preocupam em abordar o tema. Alguns gramáticos se apoiam na ideia da identificação e não identificação do elemento que ocupa a posição de sujeito. A priori poderíamos tomar como base os conceitos e torná-lo opostos, entretanto, não seria plausível uma vez que a análise de ambos vão muito além de uma mera comparação.

Milanez (1982, p.26) pontua que quando os interlocutores conseguem especificar e identificar uma das três pessoas do discurso temos um caso de determinação, diferentemente da indeterminação, em que não é possível saber cuja pessoa ocupa a posição de sujeito sintático, ou seja, podendo compreender qualquer uma delas ou as três ao mesmo tempo, sendo conhecida como a ‘não-pessoa’, havendo, portanto, uma generalização.

Em consonância com o pensamento de Milanez (1982), Cavalcante (1998) enfatiza que a referência é um fator fundamental para saber se é possível instaurar a determinação ou indeterminação, uma vez que ela pode apresentar-se específica ou não. Para a autora:

a determinação da referência às pessoas do discurso define-se, então, na medida em que aponta para as pessoas do discurso, especificando-as; enquanto a referência indeterminada é entendida a partir do pressuposto de que não é possível apontar, com precisão, as pessoas envolvidas no discurso, pois apresenta-se sob a forma genérica, podendo abranger todas as pessoas do discurso. (p. 34)

Outro conceito que é importante deixar claro é a diferença entre indeterminação e indefinição. De modo geral, a indefinição é um termo usado pela

gramática normativa para referir aos pronomes indefinidos, cuja finalidade é se referir à terceira pessoa do discurso de modo genérico, vago ou impreciso. Segundo Santana (2006, p. 43), na indefinição

[...] a referência diz respeito exclusivamente as formas lexicais de terceira pessoa: **tudo, nada, alguém, ninguém** etc., enquanto na indeterminação a generalização é essencial, podendo envolver qualquer pessoa (primeira, segunda e terceira) de forma isolada ou simultânea. (SANTANA, 2006, p. 43, grifos nosso).

Para tanto, organizamos um quadro-síntese, a seguir, com os principais critérios apresentados por Milanez (1982, p.38-42) sobre a diferença entre indeterminação e a indefinição.

Quadro 1 - Principais diferenças entre indeterminação e indefinição de acordo com Milanez (1982).

INDETERMINAÇÃO	INDEFINIÇÃO
Abrange apenas elementos humanos.	Abrange elementos humanos e não-humanos.
Apresenta recursos sintaticamente bem distintos entre si (tanto formas verbais como itens lexicais de 1ª, 2ª e 3ª pessoa).	Envolve formas lexicais exclusivas de 3ª pessoa: alguém, algo, uns, tudo, nada, etc.
Os conteúdos dos recursos vão além das marcas sintáticas, uma vez que é possível ao falante se referir as três pessoas do discurso mesmo através de um item lexical de 1ª, 2ª e 3ª pessoa.	Seus recursos são, por si, insuficientes para remeter a outras pessoas que não a 3ª.
A generalização é uma característica essencial para a indeterminação.	A generalização pode ser considerada uma possibilidade para indefinição.
Os recursos de indeterminação não apresentam aspectos quantitativos em relação a um conjunto, mesmo quando ocorrem nos contextos com referência explícitas a um grupo social.	A indefinição possui natureza limitada da generalização.
As marcas de singular e plural ficam praticamente neutralizadas nas construções indeterminadas, ou seja, o que importa é ter uma referência genérica, indistinta e não-limitada que independe de marca morfológica.	Os recursos de indefinição sempre pressupõem um conjunto fechado de elementos que pode ser expresso, sendo eles de totalidade (<i>tudo, todos</i>), esvaziamento (<i>nenhum, nada</i>) ou parcialidade (<i>alguns, uns</i>).
Para analisar sintaticamente a indeterminação, é necessário o contexto para sua interpretação.	As marcas de singular e plural são distintas no processo de definição/indefinição.

A indeterminação se situa num plano mais genérico, não-limitado que se refira às três pessoas do modo indistinto.	As indefinições podem ser analisadas dentro dos limites frasais.
---	--

Fonte: Adaptado de Milanez (1982) e de Carvalho (2010).

A respeito do uso dos artigos *o/um*, Milanez (1984) salienta que são recursos fundamentais para estabelecer a diferença entre definido e indefinido. Em se tratando da indeterminação, a autora faz uma observação importante: em contextos como em *o/um indivíduo*, *o/um cara*, por exemplo, os artigos que acompanham esses substantivos são irrelevantes, uma vez que não altera o conteúdo indeterminador.

Sobre isso, Menon (2006 [1994]) se manifesta de forma contrária ao pensamento de Milanez (1982), ao declarar que há sim uma alteração no significado no uso do artigo indefinido como elemento indeterminador. De maneira didática, podemos exemplificar assim: *a gente, a pessoa, uma gente, uma pessoa*. Menon justifica que, em *uma gente, uma pessoa*, o que perdura é o significado das palavras *gente* e *pessoa*, diferentemente das formas *a gente* e *a pessoa*, uma vez que elas são formas usadas para indeterminar a referência do sujeito. Menon justifica ainda seu posicionamento afirmando que é possível substituir o artigo indefinido por um determinante ou quantificador, no entanto esse tipo de substituição não é possível em forma nominais que ocupam posição de sujeito, visto que se trata de uma *forma cristalizada*.

Milanez (1984) salienta que por muito tempo o conceito de indeterminação, indefinição e impessoalização foi sendo confundido pelas GTs, no entanto apenas em 1957 que o Ministerio da Educação e Cultura sugeriu a modificação desse conceito a Nova Nomenclatura Gramatical Brasileira, assim alguns autores passaram a distinguir os conceitos de indeterminação e impessoalização. Com isso, foi possível ver nas gramáticas elucidções como as que encontramos em Cunha e Cintra (2001 [1985]):

Não se deve ser confundido o sujeito indeterminado, que existe, mas que não pode ou não se deseja identificar, como a inexistência de sujeito. Em ‘Chove’, ‘Anoitece’ ‘Faz frio’, interessa-nos o processo verbal em si, pois não o atribuímos a nenhum ser. Dize-se então que o verbo é IMPESSOAL; e o sujeito INEXISTENTE (CUNHA E CINTRA, 2001 [1982], p. 143).

De modo geral, notamos que Cunha e Cintra (2001 [1985]) tentam estabelecer uma distinção entre impessoal e inexistente, deixando evidente que no primeiro caso é existente a presença do sujeito, mas que pode ou não ser identificado, já no segundo, refere-se, de modo geral, aos verbos e o lugar sintático ocupado pelo sujeito na oração é

inexistente. Milanez (1982) reforça dizendo que o sujeito indeterminado é existente, ou seja, há um elemento semântico presente na oração, contudo este elemento não é representado sintaticamente nas duas ocorrências na qual a gramática normativa prevê sobre a indeterminação: verbo na terceira pessoa do plural sem sujeito explícito e verbo e verbo na terceira pessoa do singular mais o pronome *se*. Contudo, Milanez (1982) enfatiza o seguinte:

“[...] a impessoalização é uma propriedade de certos verbos que, por sua própria natureza, não atribuem a seus SN-sujeitos nenhuma função referencial. A indeterminação, ao contrário, sempre envolve uma referência. Justamente é esta ausência que diferencia as construções impessoais de todas as outras formas verbais da língua.” (MILANEZ, 1982, p. 84)

A respeito da impessoalização, a autora reforça ainda dizendo que em outras línguas como o inglês e o francês há possibilidade de preencher lexicalmente a posição do sintagma nominal, enquanto no português brasileiro não se admitem, à esquerda dos verbos impessoais, nenhum item lexical, como sinalizado no exemplo a seguir:

Ex. (35)

“It rains.” / “Chove.”
(MILANEZ, 1982, 45)

Ex. (36)

Il pleut. / “Chove.”
(MILANEZ, 1982, 45)

É com base nesse ponto de observação que a indeterminação e impessoalização se difere, uma vez que a indeterminação é um fenômeno que pode ocorrer com sujeito lexicalizado ou não, sempre envolvendo uma referência. Além disso, para que a indeterminação seja analisada, precisamos ter acesso ao contexto a qual está inserido.

Síntese Conclusiva

No presente capítulo, apresentamos pressupostos teóricos que norteiam a pesquisa no âmbito da Indeterminação da referência do sujeito, principalmente no que tange as definições de *determinação*, *indeterminação*, *indefinição* e *impessoalização*. Milanez (1982), Menon (1998) e Carvalho (2010) fundamentaram as ideias desenvolvidas ao longo do capítulo. Tomamos em particular Menon (2006 [1998]) para conceituar a indeterminação do sujeito como sendo os casos em que não se quer ou não se pode identificar o elemento que ocupa o lugar de sujeito. Com base nessa definição foi possível ter uma maior visão do objeto em estudo.

Ainda segundo a visão da autora, é possível através do contexto recuperar a qualquer momento o referente mencionado pelo locutor, o que não seria viável às formas indefinidas “um” e “outros”, por exemplo, uma vez que qualquer um que compõe o discurso pode ser o referente, inclusive o seu representante.

Para fundamentar teoricamente a respeito das diferenças entre *determinação*, *indeterminação*, *indefinição* e *impessoalização* utilizamos Milanez (1982). A autora enfatiza que embora esses conceitos aparentam ter semelhanças, é necessário que se estabeleça uma distinção, por exemplo: a determinação/indeterminação pertence ao nível da determinação, pois é possível identificar, de modo preciso e indistinto, a(s) pessoas do discurso envolvida, já a indefinição e a impessoalização não pertence a esse universo.

Outro conceito que também foi apresentamos neste capítulo foi o de sujeito gramatical. Explicamos que as definições apresentadas pelas gramáticas normativas são equivocadas, uma vez que há mistura de critérios semântico e sintático. Além disso, deve-se levar em consideração cada elemento que compõe a estrutura oracional e como eles estabelecem relação entre si.

Sobre a indeterminação, muitas pesquisas de natureza sociolinguística vêm explorando constantemente os conceitos aqui apresentado, quer seja usando dados orais, quer seja dados escritos. Assim, visando dar maior subsídio para nossa análise, apresentaremos, no capítulo seguinte, a metodologia utilizada para desenvolver nosso estudo. Para tantos especificamos as etapas trabalhadas as hipóteses levantadas e trato quantitativo dos dados.

3. PROCEDIMENTOS METODOLÓGICOS

O presente estudo ancora-se nos preceitos teórico-metodológicos da Dialetoлогия Pluridimensional (CARDOSO; FERREIRA, 1994; RADTKE; THUN, 1991, THUN, 1998; CARDOSO, 2010) e na Sociolinguística Laboviana (TARALLO, 2006; CEZARIO; VOTRE, 2008; LABOV, 2008) que considera a língua como um fato social, isto é um sistema heterogêneo marcado por variação linguística e em alguns casos ocasionando a mudança linguística.

Neste capítulo, explicitamos os passos seguidos para a concretização da pesquisa, apresentando uma breve síntese do Projeto ALiB, dando ênfase aos aspectos que são fundamentais para o conhecimento do corpus de nosso estudo. Assim, estruturamos o presente capítulo da seguinte forma: (i) perfil dos informantes; (ii) a seleção e constituição do corpus da pesquisa; (iii) um breve histórico das localidades investigadas; além do processamento estatístico dos dados a partir do Programa Goldvarb X e (iv) as variáveis analisadas.

3.1 Projeto Atlas Linguístico do Brasil - ALiB

O interesse em elaborar um atlas linguístico brasileiro surgiu em 1952, por meio do Decreto 30.643, de 20 de março, na qual a Comissão de Filologia da Casa de Rui Barbosa teve como principal finalidade a “elaboração do Atlas Linguístico do Brasil”. No entanto, apenas em 1996 esse desejo começou a ser concretizado por uma equipe de professores/pesquisadores que viram na criação de um atlas a oportunidade de coletar, analisar e descrever as particularidades do português brasileiro.

O ALiB é um projeto de abrangência nacional que objetiva fazer o mapeamento do português falado no Brasil, considerando os diferentes níveis da língua: fonético-fonológico, morfossintático, semântico lexical e prosódico. Seu interesse se pauta em descrever a realidade linguística numa perspectiva diatópica, considerando a diversidade linguística existente em cada região. Assim, sua contribuição se apoia na ideia de que não existe uma língua unificada, e que ela funciona como um instrumento de comunicação social, dotado de normas de uso, mas suscetível de variação no espaço geográfico e social, ou seja, dotada de uma heterogeneidade ordenada (CARDOSO, 2010).

Sendo assim, os dois primeiros volumes, contendo a introdução e a apresentação de cartas linguísticas do Atlas Linguístico do Brasil, foi lançado em 2014

durante III Congresso de Dialetoologia e Sociolinguística (III CIDS), realizado em Londrina, em homenagem às professoras Suzana Cardoso (*in memoriam*⁸) e Jacyra Mota, atual diretora presidente. O volume I é o de Introdução e o Volume II apresenta 159 cartas linguísticas, com dados de 25 capitais do estado.

O acervo do ALiB é formado por gravações de entrevistas realizadas com 1.100 falantes, distribuídos por 250 localidades em todo o país. Para a seleção dessas localidades foram considerados os critérios demográficos, históricos e culturais.

A respeito dos informantes, preferencialmente, foram considerados falantes nativos pertencentes às localidades, distribuídos por sexo, faixa etária e escolaridade, como especificado no subtópicos a seguir.

No que tange aos instrumentos de pesquisa usados para coleta dos dados que compõem o corpus do ALiB estão explicitados no subtópicos 3.3.

3.2 Perfil dos Informantes

O ALiB considera dados de fala de oito informantes por capital investigada, obedecendo dois níveis de escolaridade. Os informantes do ALiB foram selecionados com base no perfil descrito a seguir: sujeitos de ambos os sexos, distribuídos, igualmente, em duas faixas etárias – faixa I, de 18 a 30 anos, e faixa II, de 50 a 65 anos – e dois níveis de escolaridade – Ensino Fundamental incompleto, com, no máximo, até o 7º ano, e Ensino Universitário Completo. Eles devem ser naturais da localidade pesquisada, não tendo dela se afastado por mais de um terço de suas vidas, e seus pais devem ser, preferencialmente, naturais da mesma localidade.

Para preservar a identidade dos sujeitos da pesquisa, o ALiB os codifica da seguinte forma: aos informantes do sexo masculino atribuem-se números ímpares e, aos do sexo feminino, números pares; os números de 1 a 4 correspondem aos sujeitos que cursaram ou cursam o Ensino Fundamental, e os de 5 a 8, aos sujeitos com formação universitária. Com relação à faixa etária, os números 1, 2, 5 e 6 correspondem aos sujeitos mais jovens (faixa etária I), e os números 3, 4, 7 e 8, aos mais idosos (faixa etária II).

Para codificação das localidades, foi atribuído um número a cada uma delas, que é antecedido pela sigla do Estado. Em se tratando dos municípios *locus* da pesquisa,

⁸ Professora Suzana Cardoso é referência no que tange aos estudos dialetológicos. É inegável a sua dedicação em tornar possível e acessível o conhecimento sobre o português brasileiro. Como grande mentora, deixou como legado valioso para muitos pesquisadores.

estes estão assim identificados: São Luís – MA026, Teresina – PI034, Fortaleza – CE041, Natal – RN053, João Pessoa PB061, Recife PE070, Maceió AL077, Aracajú SE079, Salvador BA093. Para este estudo, foram selecionados 72 sujeitos, distribuídos equitativamente entre as nove localidades como mostra o Quadro 2, a seguir.

Quadro 2 - Distribuição dos sujeitos da pesquisa, segundo o perfil do ALiB.

Localidade	São Luís / Teresina / Fortaleza Natal / João Pessoa / Recife Maceió / Aracajú / Salvador			
	Ensino Fundamental		Ensino Universitário	
Sexo	Masculino	Feminino	Masculino	Feminino
Faixa etária I (18 a 30 anos)	Inf. 1	Inf. 2	Inf. 5	Inf. 6
Faixa etária II (50 a 65 anos)	Inf. 3	Inf. 4	Inf. 7	Inf. 8

Fonte: Elaborado pela autora.

3.3 Instrumentos da pesquisa

Os instrumentos de pesquisa adotados pelo ALiB são compostos por um conjunto de questionários que investiga a língua em diferentes níveis de análise linguística. Sua primeira versão foi publicada em 1998, com o objetivo de conhecer e testar esse instrumento, visando, assim, aplicar de forma experimental em diferentes pontos do país. Com base em uma análise detalhada da aplicação dos testes, os questionários foram reformulados com vista à elaboração da versão final a ser aplicada e publicada, pela Universidade Estadual de Londrina, em 2001. Convém dizer que aplicação desse questionário foi realizada por uma equipe de professores pesquisadores e auxiliares de pesquisa. Os inquéritos foram realizados nos 250 pontos linguísticos, com informantes previamente selecionados e que se encaixavam no perfil adotado pelo ALiB.

O conjunto de questionários linguísticos adotado pelo ALiB é composto por três questionários com perguntas direcionadas a diferentes fenômenos linguísticos presentes na língua falada. São, pois, eles: (i) Questionário Fonético-Fonológico (QFF), com 159 perguntas, na qual temos 11 questões de prosódia; (ii) Questionário Semântico-Lexical (QSL), 202 perguntas; e (iii) Questionário Morfossintático (QMS), 49 perguntas. Além disso, há quatro Questões de Pragmática (QP), seis Perguntas

Metalinguística (PM). Propõem-se também Temas para Discurso Semidirigido (TDS) – registros de narrativas pessoais –, além da leitura de um texto – a “Parábola dos sete vimes”.

Além dos questionários, o ALiB conta com a Ficha de Localidade e a Ficha do Informante. A primeira objetiva registrar informações geográficas, históricas e sociais da localidade investigada. É preenchida antes da entrevista e pode ser usada para subsidiar a análise dos dados. A segunda reúne informações gerais acerca dos falantes, como sexo, faixa etária, nível de escolaridade, diversões e lazer.

Após a gravação da entrevista, os áudios são salvos no computador, para serem faixalizados e salvos em CDs, para, finalmente serem transcritos.

3.4 Região Nordeste: *Loci da Pesquisa*

O Brasil possui cinco divisões regionais, sendo elas: Região Norte, Nordeste, Sudeste, Sul e Centro-Oeste, de acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE, 1990), tal divisão foi estabelecida conforme as características culturais, econômicas, sociais, políticas coexistentes no Território Nacional.

Em se tratando da Região Nordeste, sua extensão territorial equivale a 18,27% do território nacional, compreendendo uma área de 1.561.177,80 km² (ARAÚJO, 2011) e, segundo dados do IBGE (2018), conta com aproximadamente 56 milhões de habitantes. É a região geográfica do Brasil com maior número de estados, 09; sendo, pois, eles: Maranhão, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe e Bahia.

O Nordeste é conhecido pela beleza de suas praias, que correspondem a uma extensão de aproximadamente 3.338km, mas também como uma região de clima seco. Suas características físico-geográficas são formadas pelo contraste de sua paisagem, relevo, hidrografia e clima, o que condiciona a população nordestina a viver principalmente de atividades econômicas referentes ao campo, como agricultura e pecuária e o turismo (ARAÚJO, 2011).

A população nordestina caracteriza-se pela mistura de etnias, sendo estas representadas por portugueses (colonizadores), africanas (escravos) e indígenas (nativos). Em meio a esse universo de diversidade étnica, surge a figura estereotípica do nordestino associado ao sertanejo, ou seja, um homem guerreiro capaz de enfrentar as dificuldades e de sobreviver a elas. Para Albuquerque Júnior (2009):

“O tipo nordestino vai se definindo como um tipo tradicional, voltado para a preservação de um passado regional que estaria desaparecendo...se situa na contramão do mundo moderno, rejeita as suas superficialidades, sua vida delicada e histérica. Um homem de costumes conservadores, rústicos, ásperos, masculinos; um macho capaz de resgatar aquele patriarcalismo em crise; um ser viril, capaz de retirar a sua região da situação de passividade e subserviência em que se encontrava.” (ALBUQUERQUE JÚNIOR, 2009, p. 150)

Na esfera cultural, o retrato do Nordeste, dentro e fora da região, está cada vez mais marcado pelas singularidades na produção cultural, sendo possível identificar manifestações como maracatu⁹, bumba-meu-boi¹⁰, reisado¹¹, coco de roda¹², frevo¹³, literatura de cordel¹⁴, xilogravuras¹⁵, entre outras.

A língua falada é um dos elementos culturais que materializa a identidade regional, registrando e expressando as experiências e os sentimentos da vida sertaneja. Carvalho (2002, p. 50) salienta que a língua “guarda consigo toda a experiência histórica acumulada por um povo durante a sua existência” e por meio da língua temos uma dimensão da realidade linguística de uma comunidade de fala.

Albuquerque Júnior (2009) afirma que o Português Brasileiro (PB) teria se fragmentado em línguas regionais devido aos diversos núcleos de povoação e colonização, tendo sido o Nordeste o núcleo irradiador da diversidade linguístico-cultural. O autor também enfatiza que “estudar o dialeto do Nordeste significa retomar

⁹ O *maracatu* é um grupo carnavalesco pernambucano, com pequena orquestra de percussão, tambores, chocalhos, gonguê, (agogô dos candoblês baianos e das macumbas cariocas), percorre as ruas, cantando, dançando sem coreografia especial. (CASCUDO, 1972 p. 553)

¹⁰ O *bumba-meu-boi*, ou simplesmente *boi*, é a mais recorrente e conhecida manifestação cultural do estado do Maranhão. Um auto popular religioso, que mistura influências indígenas, europeias e africanas, simbolizando as principais etnias formadoras do povo brasileiro, num relato popular e dramatizado das relações sociais entre patrão e subalternos – para celebrar a fé e a devoção aos santos do catolicismo, em especial São João, Santo Antônio, São Pedro e São Marçal. Envolve em sua realização artes visuais, como os bordados nas indumentárias e no couro do boi, feitos com paetês, canutilhos e contas de vidro ou miçangas, além de teatralidade, música e dança. (FAPEMA)

¹¹ Reisado é a denominação erudita para os grupos que cantam e dançam na véspera do dia de Reis (6 de janeiro). (CASCUDO, 1972, p.774)

¹² Coco de roda ou dança do coco é uma dança popular nordestina, cantada em coro o refrão que responde aos versos do tirador de coco ou coqueiro, quadras, emboladas, sextilhas e décimas. É canto-dança das praias e do sertão. A influência africana é visível, mas sabemos que a disposição coreográfica coincide com as preferências dos bailados indígenas, especialmente dos tupis das costas. (CASCUDO, 1972 p. 292)

¹³ Frevo é uma dança de rua e salão, é a grande alucinação do carnaval de Pernambuco. Trata-se de uma marcha de ritmo sincopado, obsedante, violento e frenético, que é a sua característica principal. E a multidão ondulada, nos meneios da dança, fica a ferver. E foi dessa ideia de fervura (o povo pronuncia *fervura*, *frever*, etc.), que se criou o nome *frevo*. (CASCUDO, 1972, p.414)

¹⁴ A Literatura de Cordel é produzida a partir da narrativa popular, inicialmente verbalizada e depois impressa em folhetos, os quais eram expostos para venda de forma original, dependurados em cordas ou cordéis ou expostos no chão. (RUFFINI, 2009, p.47)

¹⁵ As xilogravuras são feitas pela impressão (sobre o papel ou outro suporte) de uma matriz em madeira. Por sua vez sua aparente simplicidade, a xilografia é a mais espontânea das técnicas gráfica. (COSTELLA, 2003.p.s/n)

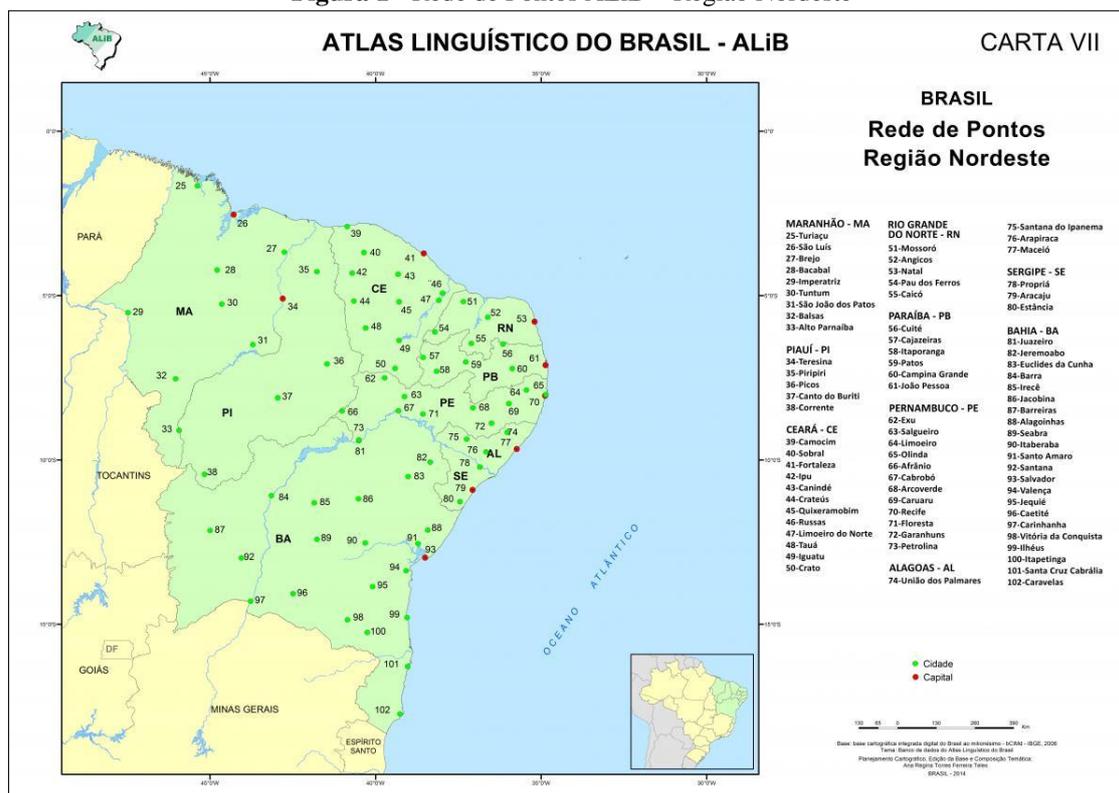
ao passado, estudar a memória de uma sociedade na qual aquelas expressões dialetais tinham se gerado” (ALBUQUERQUE JÚNIOR, 2009, p.135-136). Reforça ainda que a preocupação em relação ao falar nordestino era ir além da visão exótica e caricata, abrindo passagem para que essa variedade linguística contribuísse para um “novo falar culto”.

Assim, com base nesse pensamento, reforçamos a ideia de que a língua é um sistema heterogêneo passível às mais diversas mudanças linguísticas, seja no nível lexical, pragmático ou estrutural. Garcia (1995) ratifica pontuando que “[...] o linguajar de um vaqueiro do semiárido não é o mesmo de um trabalhador da cana ou do homem da capital. E, muitas vezes, até a mesma palavra traz outro significado quando transportada de uma área a outra.” (GARCIA, 1995, p.6)

Nesse sentido, o *loci* deste estudo foi escolhido de acordo com a rede de pontos linguísticos do Projeto ALiB (*cf.* Figura 1, a seguir), no entanto, como o Projeto conta com uma vasta rede, delimitamos apenas a capital de cada estado da Região Nordeste. Desta forma, esta pesquisa englobará todas as capitais do Nordeste, fazendo um confronto comparativo com os dados obtidos. As informações de natureza geográfica e histórica acerca dos municípios foram extraídas do site do IBGE e de obras literárias que abordam em específico cada localidade trabalhada, já as de caráter socioeconômico foram pesquisadas no site da Federação das Indústrias do Estado do Rio de Janeiro e IBGE¹⁶.

¹⁶ O IFDM – Índice FIRJAN de Desenvolvimento Municipal – é um estudo do Sistema FIRJAN que acompanha anualmente o desenvolvimento socioeconômico de todos os mais de 5 mil municípios brasileiros em três áreas de atuação: Emprego & Renda, Educação e Saúde. Criado em 2008, ele é feito, exclusivamente, com base em estatísticas públicas oficiais, disponibilizadas pelos ministérios do Trabalho, Educação e Saúde.” (Federação das Indústrias do Estado do Rio de Janeiro – Disponível em: <https://www.firjan.com.br/ifdm/>. Acesso em 17 jul. 2021).

Figura 1 - Rede de Pontos ALiB – Região Nordeste



Fonte: Projeto ALiB

3.4.1 São Luís (MA026)

São Luís, capital do estado do Maranhão, foi fundada em 8 de setembro de 1612, pelos franceses Daniel de La Touche e François de Rasilly. A capital está localizada na mesorregião Norte Maranhense, microrregião Aglomeração Urbana de São Luís. Sua extensão territorial é de aproximadamente 834,785 km², com a estimativa de 1.014.837 habitantes. Conhecida pelo povo Tupinambá como a Ilha de Upaon-Açu ou Grande Ilha, São Luís representa um dos principais polos urbanos do Estado.

A capital maranhense reúne diferentes culturas, refletindo em seu patrimônio arquitetônico a influência portuguesa. Na obra *Poranduba Maranhense (1819/1820)*, de Frei Francisco de Nossa Senhora dos Prazeres Maranhão, o autor faz um dos primeiros registros sobre as fortes contribuições de povos ameríndios e africanos, sobretudo no português maranhense.

Em se tratando do Censo demográfico realizado em 2010 pelo IBGE, cerca de 94% dos moradores vivem na área metropolitana de São Luís. Convém dizer que aproximadamente 20% são de migrantes oriundos de interiores do Estado, assim como de outros estados, na busca de oportunidades de vida. De acordo com as informações

disponibilizadas pelo o Firjan (2015), a Capital tem como principais atividades econômicas a agricultura, a pecuária e o extrativismo vegetal.

3.4.2 Teresina (PI034)

A capital do Piauí, Teresina, está localizada no Centro-Norte do Estado, a 366 km do litoral. Segundo dados do IBGE (2020), a cidade possui 1.392 km² de extensão territorial e 868.075 mil habitantes. Começou a ser povoada no século XVII, por Domingos Jorge Velho e um grupo de bandeirantes. Oficialmente sua fundação ocorreu em 16 de agosto de 1852, após a construção da igreja de Nossa Senhora do Amparo. Convém dizer que Teresina foi construída durante o reinado de D. Pedro II, sendo considerada a primeira capital brasileira planejada. Segundo relatos históricos, durante o período colonial destaca-se o forte apoio da Imperatriz do Brasil Dona Teresa Cristina Maria de Bourbon as tomadas de decisões junto ao Imperador, o que acabou refletindo na denominação da cidade com um topônimo que até hoje se conserva: “**Teres**, tirado de Teresa, e **ina** de Cristina, Teresina, Cidade Verde, conforme a cognominou Coelho Neto” (ROMÃO DA SILVA, 1994, p. 19).

Assim como em São Luís, Teresina também sofreu fortes influências dos colonizadores, africanos e indígenas, principalmente no falar, que vai além das barreiras, sendo observadas nos municípios vizinhos. Além disso, os colonizadores paulistas também contribuíram para o linguajar falado em Teresina, na época, recebendo a denominação de “arraiá dos paulistas” local onde as fazendas cultivavam a terra e criavam gado. Essa denominação ocorreu pelo fato de os primeiros moradores serem de São Paulo. (NUNES, 1975).

No que tange ao aspecto econômico, a maioria do desenvolvimento econômico de Teresina tem forte sustentação no setor terciário, ou seja, área comercial, serviços autônomos e administração pública. Os setores de bebidas, alimentícios, de bicicletas, colchões, metalúrgicos e vestuários são os que mais se destacam na Capital. No entanto, a agricultura e a pecuária também tem fundamental importância, principalmente com a produção de cana-de-açúcar, aves e rebanho bovino.

3.4.3 Fortaleza (CE041)

Fortaleza, capital do Ceará possui um território que compreende a extensão de 312.353km² e atualmente conta com uma população estimada de 2.703.391 habitantes. É

a capital brasileira geograficamente mais próxima da Europa, em particular da cidade de Lisboa, Portugal. Por conta disso, Fortaleza serviu como principal ponto de transferência (entrada e saída) entre o Brasil e os quatros continentes. (BARBOZA, 2013)

A cidade surgiu ao redor do Forte de Nossa Senhora da Assunção, construído pelos portugueses, em 13 de abril de 1726, deixando de ser povoado e passando para a categoria de vila. No entanto, somente em 1823 o Imperador Dom Pedro I elevou a vila à categoria de cidade.

Na época, a produção e comércio de algodão foram às bases da economia cearense, promovendo o comercial e político da capital. Atualmente, suas maiores atividades econômicas são a pecuária, com o rebanho efetivo na área de pecuária leiteira. Na agricultura podemos destacar a produção de grãos como feijão e milho. O turismo também impulsiona a economia da cidade, visto que conta com belíssimas praias tais como Iracema, Meireles e Mucuripe.

Em relação ao acervo linguístico presente em Fortaleza, notamos que é bastante rico e vasto, com fortes influências de diferentes grupos indígenas que viviam principalmente na região litorânea. Essa contribuição está presente nos topônimos, como *Guararapes*, nomes da botânica e da zoologia regional, como *japicanga*¹⁷ e *aruá*¹⁸, respectivamente. (FILHO, 1962)

3.4.4 Natal (RN053)

Capital do Rio Grande do Norte, Natal possui área de 167 km² e população estimada de 896.708 habitantes. Sua fundação ocorreu em 25 de dezembro de 1599, às margens do Rio Ponteji. Na época, o Brasil estava sob domínio espanhol, no entanto, antes de fundar a capital, os colonos agruparam-se próximos ao “Forte dos Santos Reis” para se defender dos nativos que habitavam na redondeza, dando origem mais tarde a chamada de “Cidade dos Reis”.

Nessa época, era significativa a presença dos indígenas, com os Tupis e Cariris, negros, que vieram da Guiné e os colonizadores, que era composto por portugueses e holandeses. (CASCUDO, 1984). A presença desses povos reflete na variação linguística

¹⁷ *Japicanga* é uma chibata de um cipó flexível cheio de nozinho, assim chamada do nome vulgar da planta que é tirada, a *Smilax japicanga*, da família das asparagáceas, de virtudes medicinais, e que vegeta às margens dos rios e em lugares frescos. (CASCUDO, 1972, p. 472)

¹⁸ *Aruá* é um molusco gastrópode, gênero ampulária, uruá. “Pequeno molusco gastrópode, espécie de búzio ou concha, espiral, ovoide, de um verde escuro, quase preto, que se encontra nos alagadiços, água doce, contendo uma espécie de marisco que se come guisado.” (CASCUDO, 1972, p. 107)

da região, em especial no léxico, como é o caso da palavra *beliro* usada para denominar grampo de cabelo (NASCIMENTO & PAIM, 2020)

Após a colonização, foi preciso desenvolver economicamente a cidade, por conta disso, essas atividades começaram pelo litoral, visto que era caminho para as rotas marítimas que ali passavam e atracavam. Nos dias atuais, as principais atividades econômicas desenvolvidas são a produção de camarão, galinhas e caprinos. Na agricultura, ganha destaque a cultura de banana, laranja, manga e castanha de caju. Assim como Fortaleza, Natal também se mantém do turismo, visitas às praias de Ponta Negra, o Morro do Careca e a Fortaleza dos Reis Magos.

3.4.5 João Pessoa (PB061)

Também conhecida como “Porta do Sol”, João Pessoa, capital da Paraíba está localizada na Ponta da Seixas, local onde “o sol nasce primeiro no continente americano”. A cidade possui área de 210,044 km², com o número atual de aproximadamente 825.796 habitantes. Sua fundação aconteceu em 5 de agosto de 1585 e recebeu inicialmente o nome de *Cidade Real de Nossa Senhora das Neves*. João Pessoa recebe o título de terceira cidade mais antiga do Brasil, tendo sido fundada pela Cúpula da Fazenda Real, já sendo considerada cidade e não como vila, povoado ou aldeia.

Assim como as outras capitais do Nordeste, João Pessoa também recebeu fortes influências dos povos indígenas na sua formação. Relatos históricos narram que era significativa a presença dos povos potiguaras e tabajaras. O cronista Pero de Magalhães Gandavo (séc XVI), comenta no trecho a seguir que a língua dos índios:“(…) carece de tres letras, convem a saber, nam se acha nella F nem L, nem R, cousa digna despanto porque assi nam têm Fé, nem Lei, nem Rei, e desta maneira vivem desordenadamente sem terem alem disto conta, nem peso, nem medido.” (GURJÃO *et. al.* 2001, p. 182, *apud* GANDAVO, 1980: 54).

Sobre a economia, João Pessoa também investiu na produção de camarão, assim como em Natal. Além disso, a cidade se destaca pela produção de ovos de codornas e da pecuária leiteira. Na agricultura, damos destaque para o cultivo de cocoda-baía, maracujá e manga. No entanto, o turismo é o que mais produz renda e gera empregos. Convém dizer que outra atividade econômica que vem ganhando destaque é a intensa expansão imobiliária que tem mostrado que esse tipo de empreendimento está impulsionando cada vez mais a economia da capital.

3.4.6. Recife (PE070)

Recife é a capital do Estado de Pernambuco, com cerca de 218 km² em extensão territorial e população estimada de 1.653.461 habitantes. A capital surgiu no ano de 1537 como “Ribeira de Mar dos Arrecifes dos Navios”, por conta disso a atividade comercial desenvolveu-se rapidamente impulsionando o crescimento do povoado, logo ficou conhecida devido à cultura da cana-de-açúcar e ao pau-brasil (ou pau-de-pernambuco).

Silva (2001) salienta que o vocabulário encontrado no falar de Recife se fazia presente nas cartas e mapas dos portugueses, sendo possível ver em sua forma primitiva, como *Perñabuquo* e *Fernambouc*, ambas originárias da forma tupi (Paraná-puka) que significa mar furado ou buraco de mar. Essa observação nos revela as fortes influências dos povos indígenas na região onde hoje se encontra a capital de Pernambuco, Recife.

A economia de Recife se alicerça, assim como Natal e João Pessoa, na produção de camarão e da pecuária leiteira. Damos destaque também à produção de suínos e galinhas. No entanto, o que mais se sobressai são as atividades voltadas para o setor terciário (serviços, administração pública e comércio), assim como setor da construção civil e da indústria de transformação. O turismo também se tornou forte aliada à economia da capital, isso porque a cidade possui vários pontos turísticos e culturais, por exemplo: Atelier de artes cerâmicas de Francisco Brennand, Bairro do Recife Antigo e Praia de Boa Viagem.

3.4.7. Maceió (AL077)

A capital do Estado de Alagoas, Maceió, possui área de 509.552 km², com aproximadamente 1025.360 habitantes. O povoado que deu origem a Maceió surgiu em um engenho de açúcar, no ano de 1609. Nessa época, os índios tapuias que ali habitavam foram expulsos para a Amazônia por colonizadores portugueses, que logo construíram o Porto de Jaraguá para assegurar o desenvolvimento econômico da cidade na época, que se preocupava em produzir e exportar cana-de-açúcar.

É significativa a contribuição dos povos indígena na formação do léxico de Maceió, principalmente em se tratando dos estudos onomásticos, em especial a toponímia. Esse acervo lexical toponímico deve-se particularmente ao início da formação da cidade, visto que na época de colonização era significativa a presença do

povo tapuia. Bairros cujos nomes são *Jaraguá* e *Mangabeira* são exemplos de nomes com etimologia indígena. (SILVA; MELO, 2014)

Sobre a economia, Maceió expandiu seus horizontes. Atualmente, a cidade tornou-se o maior produtor brasileiro de sal-gema. Novas indústrias foram surgindo, tais como: indústrias químicas, açucareiras e de álcool, de cimento e alimentícias. Na pecuária destacamos a pecuária leiteira, a produção de ovos de codornas, tambaqui e tilápia. No setor agrícola, temos, assim como em Natal, a produção de banana, coco-da-baía e manga. Além disso, a extração de gás natural e petróleo também vêm impulsionando o setor econômico, assim como o turismo.

3.4.8. Aracaju (SE079)

A fundação da cidade de Aracaju, capital de Sergipe, está intimamente ligada à cidade de São Cristóvão, que na época era a capital da Capitania de Sergipe. Como São Cristóvão já não oferecia mais garantias nem condições para ser sede administrativa foi necessário criar um porto que garantisse a escoação da produção. Por necessidade econômica, Aracaju foi fundada em 1855, sendo a segunda capital planejada de um estado brasileiro. O projeto das ruas seguiu o modelo de um tabuleiro de xadrez para que fosse possível desembocar no Rio Sergipe, que na época favorecia o escoamento da produção açucareira. Sua extensão territorial é de 181.857 km² com população de aproximadamente 664.908 habitantes.

Durante o período da colonização, Aracaju recebeu influência dos indígenas, africanos, portugueses e holandeses que contribuíram para a formação lexical desse lugar. No entanto, Aracaju, assim como outras capitais do Nordeste, também apresenta suas particularidade e peculiaridade em relação às expressões. O Jornalista Rubens Dórias publicou no ano 2000 o Dicionário Sergipanês, que nos releva algumas das expressões típica desse lugar, por exemplo, “a pulso” que significa ‘à força’, “goelar” usado para quem pratica o ato de furtar, dentre outras. (PREFEITURA DE SERGIPE, 2010)

A respeito da atividade econômica atual, ressaltamos a criação de camarão, da pecuária leiteira e de um efetivo rebanho de equinos. Na agricultura, o coco-da-baía representa o produto mais cultivado na capital. Entretanto, a atividades de comércio, prestação de serviços, industriais e o turismo são os pilares da economia do povo aracajuano.

3.4.9 Salvador (BA093)

Popularmente conhecida como “Bahia de Todos os Santos”, Salvador, capital da Bahia, possui área de estimada de 693,8 km² e aproximadamente 2.886.698 habitantes. Seus primeiros habitantes foram os índios tapuias, que por volta de 1000 anos foram expulsos pelos tupinambás. Por volta de 1510, a embarcação que trouxe a tripulação de Diogo Álvares, também conhecido como Caramuru naufragou na região do Rio Vermelho. Salvador foi fundada em 1549 pelo Governador-Geral do Brasil Tomé de Sousa e se tornou a primeira capital do Brasil (até 1763). Em sua tripulação, o então Governador-Geral trouxe militares, médicos, farmacêuticos e fidalgos, além de padres jesuítas. Sua principal atividade econômica na época era o cultivo e produção de cana-de-açúcar.

Salvador é considerada centro da cultura africana. A cidade baiana exhibe a beleza do negro, deixando evidentes suas vestimentas, artesanatos, rituais religiosos e culinária. Melo (1981, p. 76) salienta que após o estabelecimento da população escrava no Brasil duas ‘línguas gerais’ foram constituídas: “nagô ou iorubá na Bahia e o quimbundo nas outras regiões”. Os africanos que vieram para o trabalho forçado no período de colonização do Brasil trouxeram consigo sua religião conhecida como candomblé. As festas realizadas por eles obedecem a ciclos e comemorações a Oxalá¹⁹.

Em se tratando da economia de Salvador, atualmente ela gira em torno dos centros comerciais, como o Shopping da Bahia, o Salvador Shopping e o Shopping Bela Vista. Na construção civil temos as grandes construtoras responsáveis pelas mais diversificadas edificações. As indústrias da região são de móveis, confecções, madeira, alimentos, papel, materiais elétricos; já a atividade turística e cultural representa importantes geradores de emprego e renda, sobretudo a preservação do patrimônio artístico e cultural.

3.5 O corpus e o tratamento quantitativo

O *corpus* de análise desta pesquisa foi selecionado com base nos dados de fala de entrevistas que compõem o banco de dados do ALiB. Tendo em vista que este trabalho é um estudo sobre as formas nominais de referência sujeito, variante de

¹⁹ Oxalá é o maior dos orixás, entidade andrógina, é a maior tradição religiosa como sobrevivência africana na Bahia. “Orixalá ou Oxalá” tem um caráter bissexual e simboliza as energias produtivas da natureza. [...] Pela conveniência e força da aculturação, Oxalá identificou-se com o mais popular e prestigioso culto de toda a Bahia, Nosso Senhor do Bomfim. (CASCUDO, 1972, p. 651)

natureza morfossintática, consideramos os dados coletados em todo o inquérito, visto que somente os Discursos Semidirigidos não nos forneceram dados suficientes para a pesquisa, além de verificar como esse fenômeno se comporta em contextos espontâneos de fala.

Para tornar didática a visualização dos dados, e, conseqüentemente, facilitar a análise, optamos por organizar o *corpus* em uma planilha do Excel, dividindo-a em oito colunas. Assim definimos que a primeira coluna corresponde ao amálgama dos códigos atribuídos às variáveis linguísticas e extralinguísticas, seguida da variável dependente. As colunas de três a seis dizem respeito aos fatores sociais – *localidade, sexo, faixa etária e escolaridade*. Já as colunas oito e nove correspondem aos fatores linguísticos – *tempos verbais, seguido do grau de indeterminação do sujeito*. A última coluna, nomeada de Contexto, refere-se aos trechos retirados das transcrições em que ocorreu o uso das formas nominais, como demonstrado na Figura 2.

Figura 2 - Planilha dos dados

A1	Variável Dependente	Localidade	Sexo	Faixa Etária	Escolaridade	Tempos e modos verbais	Grau de indeterminação	Contexto
(l)ãhpf	1	A	h	p	f	n	j	INQ.- Aqui, né? Mas qual é, digamos assim... eh... geralmente essas casas são como aqui? Fala um pouco dessas casas! INF.- As casas, eh... sala, quarto, cozinha, banheiro. O pessoal dá mais preferência, né, hoje aqui. Assim, bairros também mais sossegados, assim, mais pro lado de condomínios. Vinhais, São Francisco... é que o pessoal dá mais preferência.
(l)ãhpf	1	A	h	p	f	n	j	INQ.- Aqui, né? Mas qual é, digamos assim... eh... geralmente essas casas são como aqui? Fala um pouco dessas casas! INF.- As casas, eh... sala, quarto, cozinha, banheiro. O pessoal dá mais preferência, né, hoje aqui. Assim, bairros também mais sossegados, assim, mais pro lado de condomínios. Vinhais, São Francisco... é que o pessoal dá mais preferência. INQ.- Uma pequena grade de metal, de ferro que se coloca em cima da churrasqueira ou da brasa para assar a carne, frango, como é que a gente chama essa grade de metal? INF.- A gente chama de... que fica em cima, né? Aquela grade de metal... INQ.- É. Você bota carne e frango para assar. Na

Fonte: Elaborado pela autora.

A respeito do tratamento dos dados, escolhemos o programa *GoldVarb X*, programa computacional usado para processar grande volume de dados linguísticos. Sua primeira versão, denominada de *Varbrul (Variable rules analysis)*, foi desenvolvida no ano de 1978 por David Sankoff e Pascale Rousseau. Segundo Guy e Zilles (2007, p. 105) o programa “[...] permite investigar situações em que a variável linguística em

estudo é influenciada por vários elementos do contexto, ou seja, múltiplas variáveis independentes”.

Assim, o programa permite-nos ter acesso às informações como os percentuais e os pesos relativos referentes às variáveis independentes. Guy e Zilles (2007) salientam que:

O peso relativo de um fator é o valor calculado pelo Varbrul (com base em um conjunto de dados) que indica o efeito deste fator sobre o uso da variante investigada neste conjunto. O valor dos pesos recai sempre no intervalo entre zero e um (0-1), em que um valor de zero indica que tal variante nunca ocorre quando este fator está presente, e um valor de 1 indica que tal variante sempre ocorre quando o fator está presente[...] Os valores dos pesos recaem no mesmo intervalo entre 0 e 1 que as probabilidades, e representam tendências probabilísticas. (GUY; ZILLES, 2007, p. 239).

Nesse sentido, esperamos, por meios das rodadas estáticas realizadas no programa GoldVarb X, verificar como a nossa variável de estudo se comporta perante os fatores aqui considerados para análise.

3.6 Organização do envelope da variação

Para os estudos sociolinguísticos, o envelope de variação pode ser conceituado como a descrição detalhada das variantes, ou seja, os contextos em que as formas alternantes são empregadas na língua em uso contendo o mesmo sentido (TARALLO, 2002). Antes de se estabelecer um envelope de variação considerando equivalência semântica, observamos não só a possível variação entre as estruturas (sintáticas ou não), mas também como essas variantes se comportam no contexto em uso.

Guy e Zilles (2007) reforça esse pensamento afirmando que:

“A identificação de uma variável inclui definir as variantes (o que é e o que não é uma ocorrência da variável em estudo) e determinar o envelope da variação (onde é possível ou impossível que a variável ocorra). Contextos categóricos (nos quais não há variação) e contextos neutralizadores (nos quais a variação é irrelevante ou imperceptível) devem ser identificados e, normalmente, são excluídos da análise.” (GUY; ZILLES, 2007, p. 36)

Assim, tomando como base essas observações apresentada e a fim de tonar este trabalho mais didático, de modo a facilitar nossa análise, julgamos pertinente montar um envelope de variação. Primeiramente, escolhemos uma variável linguística de natureza morfossintática, neste caso selecionamos as formas nominais de referência do sujeito nas capitais nordestinas. Após a seleção da variável dependente, foi possível

identificar todas as formas de referenciação do sujeito utilizadas pelos informantes nordestinos analisados, conforme explicitado a seguir.

Variável Dependente: formas nominais de referenciação do sujeito

- **Grupo A:** artigo + substantivo

Ex. (37)

INQ. – Aqui, né? Mas qual é, digamos assim... eh... geralmente essas casas são como aqui? Fala um pouco dessas casas!

INF.– As casas, eh... sala, quarto, cozinha, banheiro. **O PESSOAL** dá mais preferência, né, hoje aqui. (MA26/1)

Ex. (38)

INQ. – E essa mesma carteira caiu da bolsa de uma mulher idosa e ela não viu. Como é que o rapaz jovem chama a atenção desse idoso?

INF. – **AS PESSOAS** chama de vó, de tia, ô tia sua cartêra caiu, a senhora dexô a senhora cartêra caí. (BA093/3)

- **Grupo B:** estrutura com um único elemento, ou seja, um substantivo.

Ex. (39)

INQ. – Chama de outro jeito

INF. – Chama. **NEGO** chama de sentina. Vaso sanitário. (BA093/3)

Ex. (40):

INQ.– Uma pessoa está internada em um hospital, né, e quer saber do médico se vai sair naquele dia. Como é que se pergunta? Vai perguntá ao médico, vai perguntá ao médico.

INF.– Doto, **FULANO** vai tê alta? (CE041/3)

Para o nosso estudo obtivemos duas variantes: i) a primeira diz respeito ao Grupo A, com estruturas como *O pessoal, A pessoa, As pessoas, O rapaz*, ii) a segunda variante, Grupo B, está representada por formas nominais como *Nego, Fulana(o), Pessoa*, dentre outras. Convém dizer que os dois grupos são de formas inovadoras e que a divisão dotada neste trabalho foi a melhor encontrada para apresentar os dados. Assim, com base nas variantes selecionadas, foi possível estabelecer uma análise mais significativa a respeito do fenômeno estudado. No entanto, para se analisar essas variantes, é necessário de variáveis sociais e linguísticas que ajude no melhor entendimento da variação. No tópico a seguir, apresentamos os fatores selecionados para análise.

3.7 Escolhas dos fatores de análise

É fato consabido que usar variáveis linguísticas, isto é, internas à língua, e extralinguísticas, de natureza geográfica, sociológicas e contextuais, para analisar os níveis de variações linguísticas configura-se como pontos fundamentais da Sociolinguística. É oportuno salientar que um dos objetivos das variáveis linguísticas é “[...] entender quais são os principais fatores que motivam a variação linguística, e qual a importância de cada um desses fatores na configuração do quadro que se apresenta variável” (CEZARIO; VOTRE, 2008, p.141). Ao relacionar fatores linguísticos e extralinguísticos ao fenômeno aqui estudado, pretendemos verificar até que ponto eles influenciam na escolha das variantes usadas pelo informante. Assim, elencamos a seguir, os fatores que pretendemos analisar.

Quadro 3 - Variáveis analisadas

FATORES	VARIÁVEIS
Sociais	<i>Localidade</i> (São Luís, Teresina, Fortaleza, Natal, João Pessoa, Recife, Maceió, Aracaju, Salvador)
	<i>Faixa etária</i> (Faixa I – 18 a 30 anos; Faixa II – 50 a 65)
	<i>Sexo</i> (homem e mulher)
	<i>Escolaridade</i> (ensino fundamental incompleto e ensino universitário completo)
Linguísticos	<i>Tempos e modos verbais</i> (modo indicativo e subjuntivo)
	<i>Grau de indeterminação</i> (indeterminação completa, indeterminação parcial com referência implícita no contexto, indeterminação parcial com referência explícita no contexto)

Fonte: Elaborado pela autora.

3.7.1 Fatores Sociais

Os primeiros estudos que estabeleceram uma relação entre a língua em uso e aspectos sociais foram desenvolvidos por Labov (2008), no ano de 1963, após a publicação de seu trabalho feito na comunidade de Martha's Vineyard, Massachusetts, sobre centralização do ditongo /ay/ e /aw/. Na década seguinte, Labov conclui seu outro estudo realizado na cidade de Nova York acerca da estratificação social do inglês, ou

seja, “um modelo de descrição e interpretação do fenômeno lingüístico no contexto social de comunidades urbanas” (Alkimin; 2001), conhecida como Teoria da Variação.

Assim, com base nesses estudos foi possível perceber que a concepção de língua estudada de forma isolada, sem considerar os fatores sociais, seria algo irrefutável, uma vez que o meio social em que o falante está inserido influencia nas suas escolhas linguísticas, possibilitando que a língua assuma um caráter ativo, mutável e que pode sofrer uma significativa influência de fatores como o sexo ou classe social do indivíduo, por exemplo.

Ao optarmos também pela metodologia laboviana, pretendemos apresentar de forma quantificada a influência dos fatores sociais selecionados (localidade, sexo, faixa etária, escolaridade) sob o nosso objeto de estudo, e dessa forma observar quais variáveis estão colaborando para que haja uma variação linguística no português falado nas capitais nordestinas. Para tanto, elencamos as seguintes variáveis.

a) Localidade

Com o objetivo de identificar possíveis variantes características das regiões elencada neste trabalho, bem como verificar se há diferenças em relação à frequência de uso, julgamos pertinente investigar esse fator em nossas análises. Como pontua Santana (2006, p. 92) o controle dessa variável “baseia-se na premissa de que o comportamento lingüístico de falantes de um mesmo idioma pode diferenciar-se de acordo com o maior ou menor grau de isolamento que ocupam e que determinam situações sociais específicas”.

Considerando a divisão dialetal proposta por Antenor Nascentes, resolvemos verificar se a hipótese a seguir se confirma: há uma tendência de São Luís e Salvador terem mais diferentes tipos de *formas nominais de referência* (FNs) se comparado as outras localidades. Isso porque ambas as capitais estão situadas geograficamente mais próximas de outras divisões dialetais, São Luís por se situar próximo às áreas do Falar Amazônico, e Salvador por estar próxima do Falar Fluminense, como podemos observar na figura a seguir.

Figura 3 - Divisão Dialetal proposta por Antenor Nascente



Fonte: Nascente, 1922, p. 18.

Nos exemplos a seguir, retirados do nosso *corpus*, podemos ver algumas das formas nominais encontradas em São Luís e em Salvador, nos possibilitando ter uma visão do fenômeno em estudo.

Ex. (41)

INQ.– Ah, também? Você tem alguma explicação pra essa... barriga branca?

INF.– Num sei, os... os... os... **OS COLEGAS** na faculdade chamava... falavam que era aquele marido que... que a mulhé manda nele e... tal e... e ele é corno, então chama barriga branca. (MA026/8)

Ex. (42)

INQ.– E a parte do dia quando começa a clarear?

INF.- A dependê do lugá também pode sê poluição

INQ.– Ah, é?

INF.– A poluição é forma de neblina, ela iscurece tudo, as veze **A PESSOA** até troca pura neblina, mas na verdade é poluição. (BA096/8).

b) Sexo²⁰

As diferenças entre homens e mulheres vão além do caráter fisiológico. Em se tratando dessas diferenças, podemos citar os papéis que esses agentes exercem sobre a comunidade a qual fazem parte e que refletem intimamente no seu comportamento linguístico. Labov (2008 [1972]) e outros estudiosos da área salientam que há uma maior tendência de as mulheres usarem as formas padrão da língua, ou seja, de maior prestígio do que os homens. Convém dizer que essas escolhas não estão condicionadas por fatores biológicos, mas sim por imposições sociais.

Para Fernández (1998, p. 35), “o sexo pode ser mostrado, portanto, mais como um fator de segunda ordem, como algo que normalmente subordinados a diferentes dimensões sociais e com maior poder de determinação”²¹. O comportamento linguístico de falante do sexo masculino pode ser condicionado pela localidade a qual o informante faz parte, por exemplo.

Os resultados apresentados por Santana (2006) sobre a influência da variável sexo no uso de a *gente* como recurso para indeterminar a referência do sujeito no português falado em Bananal mostra que as mulheres da faixa 1 e 2 tendem a fazer maior uso do indeterminador *a gente* e os homens, de qualquer faixa, usam mais a forma *nós*, contradizendo os postulados sociolinguísticos.

Assim, buscamos em nosso estudo analisar a distribuição das FNs entre homens e mulheres, além de verificar qual/quais são mais usadas entre os dois sexos. A hipótese é que as mulheres terão maior frequência de uso das formas nominais que os homens, visto que elas são consideradas detentoras no processo de variação linguística, mesmo a variante sendo inovadoras.

c) Faixa Etária

A faixa etária do falante configura-se como um fator significativo nos estudos labovianos, pois ela nos permite verificar se um dado linguístico está em processo de estabilidade ou mudança em progresso na comunidade de fala. Além disso, os estudos

²⁰ Não é objetivo desta pesquisa tratar das discussões existentes acerca de gênero e sexo. Para este estudo, consideramos, apenas, a nomenclatura sexo, a qual fará referência a homens e mulheres.

²¹ Tradução livre de: “El sexo puede mostrarse, por tanto, más como un factor de segundo orden, como algo que suele subordinarse a dimensiones sociales diferentes y con mayor poder de determinación.”

que envolvem a faixa etária como fator de análise nos permite comprovar a existência de variantes conservadoras ou inovadoras. Fernandez (1998) enfatiza que:

A idade dos falantes [...] é um dos fatores sociais que com maior força e clareza pode determinar os usos linguísticos de uma comunidade de fala. Dentro de certa forma, pode-se afirmar que a idade condiciona a variação linguística com mais intensidade do que outros fatores, também importante como o sexo ou classe social (FERNANDEZ, 1998, p.40)²²

A respeito da idade, Tarallo (2002) enfatiza que a faixa etária mais avançada tem maior preferência em usar as formas conservadoras, diferente dos falantes mais jovens que preferem as formas inovadoras, principalmente no nível lexical da língua, porém isso não impede a comunicação entre as faixas.

Assim, nosso objetivo é verificar em qual das faixas etárias as FNs são mais registradas. Nossa hipótese é que mesmo que as formas nominais sejam consideradas variantes inovadoras, se comparada às formas tradicionais apresentadas nas gramáticas tradicionais (a forma [ø + 3ª pessoa singular], a forma [ø + infinitivo]), a faixa etária mais avançada terá maior preferência pelas formas como *O pessoal/ A pessoa*.

Para esse estudo, distribuímos os falantes da seguinte forma:

- Faixa I – de 18 a 30 anos
- Faixa II – de 50 a 65 anos

d) Escolaridade

O fator escolaridade também foi considerado em nosso estudo, visto que em todas as línguas podemos encontrar formas linguísticas socialmente aceitas, ou seja, prestigiadas e formas que recebem estigma. Sabemos também que a escola pode ser responsável por mudanças no comportamento linguístico do falante, acelerando ou retardando esse processo.

Sobre isso, Fernández (1998) assinala que:

“A sociolinguística [...] provou que o nível de escolaridade dos falantes determina direta e claramente a variação linguística: é normal que pessoas

²² Tradução livre de: La edad de los hablantes [...] es uno de los factores sociales que con mayor fuerza e claridad pueden determinar los usos lingüísticos de una comunidad de habla. En cierto modo, puede afirmarse que que la edad condiciona la variación lingüística con más intensidad que otros factores, también importantes como el sexo o la clase social. (FERNÁNDEZ, 1998, p. 40)

mais instruídas façam maior uso das variantes consideradas mais prestigiosas ou que melhor se ajustem à norma” (FERNÁNDEZ, 1998, p.40)²³

Por essa razão, julgamos necessário analisar o efeito que a escolaridade tem sobre o uso das formas nominais de referenciação do sujeito. Como sinalizado anteriormente, os informantes desta pesquisa possuem o ensino fundamental incompleto ou universitário completo.

Ratificamos nossa escolha com os postulados de Silva e Paiva (1998, p. 342) sobre as preferências linguísticas do falante, em que diz: “[...] um padrão geral que associa a predominância às formas linguísticas padrão a falantes com maior escolarização”. Sendo assim, a hipótese a cerca da escolaridade que levantamos a respeito é que quanto maior a escolarização, mais favorável será o uso de recursos para indeterminar a referência do sujeito.

3.7.2 Fatores Linguísticos

As variáveis de natureza linguística são de suma importância para se compreender como o fenômeno linguístico se manifesta na língua. Sobre o assunto López Morales (2004) enfatiza que

“[...] todos os fatores sociais, por mais importantes que sejam, estão sujeitos aos imperativos do sistema linguístico; Até hoje, provou-se repetidamente que os fatores sociais não são tão categóricos quanto os gramaticais: eles atuam onde o sistema permite.” (LÓPEZ MORALES, 2004, p. 178-179, *apud* FERNÁNDEZ, 1998, p. 36).

Para este estudo, foram consideradas duas variáveis linguísticas: *tempos verbais* e *grau de indeterminação do sujeito*. Cada uma dessas variáveis passa a ser descrita a seguir.

a) Tempos verbais

Menon (2006 [1994]) salienta que os tempos e modos verbais possibilitam o uso da indeterminação do sujeito, em especial as construções sintáticas no *presente do indicativo*. Segundo a autora “[...] o presente tem um valor aspectual neutro, pois ele

²³ Tradução livre de: “La sociolingüística [...] há comprobado que el nivel educativo de los hablantes determina de forma directa y clara la variación linguística: es normal que las personas más instruidas hagan mayor uso de las variantes que son consideradas como más prestigiosas o que más se ajustan a la norma.”

denota uma ação sem fazer referência à sua duração e é utilizado por indicar fatos habituais” (MENON, 2006 [1994], p. 163). Com base nesta visão, resolvemos analisar esse fator e verificar a influência desta variável sobre os nossos dados, levando em consideração a classificação proposta pelas Gramáticas Tradicionais.

- Presente
- Pretérito
- Futuro
- Infinitivo
- Particípio
- Gerúndio

b) Grau de indeterminação do sujeito

O fator *Grau de indeterminação do sujeito* configura-se como sendo habitual nos estudos que investigam o fenômeno da indeterminação do sujeito, isso porque está intimamente associado aos fatores semântico-pragmáticos, ao contexto e à própria forma lexical dos recursos. Milanez (1982) enfatiza que: “[...] a indeterminação não se apresenta como um fenômeno de natureza absoluta, que se opõe à determinação sempre de forma nítida; ao contrário, a mesma comporta graus quanto à sua abrangência (MILANEZ, 1982, p. 80).”. Sendo assim, a autora traz em seu estudo três graus de análise para identificar a indeterminação do sujeito. Para efeito de análise, consideramos todos os graus estudados pela autora. Nesse sentido, os elencamos, a seguir:

- **Indeterminação completa:** quando não é possível recuperar o referente mesmo que seja apresentado o texto e/ou contexto como um todo.

Ex. (43)

INQ.– E aquele clarão que surge no céu em dias de chuva? Você já falou de raio, mas não é o raio, ele é...aliás ele é...acho que é antes do raio. É um clarão que aparece, assim, de repente.

INF.– Que **O POVO** chama relâmpago (MA026/3)

- **Indeterminação parcial com referência implícita no contexto:** quando não é possível no texto e/ou contexto identificar qualquer elemento que possa inferir o referente, a não ser pelo contexto a qual está inserido.

Ex. (44)

INQ. – Tá bom. E antigamente, será que o pessoal falava diferente?

INF. – Ah, eu acho que **O POVO** falava diferente um pouco as coisa errada.

INQ. – Eh, né?

INF. – É Eh, puquê... que tem gente antigo ainda que, que fala umas coisinhas errada. (SE079/1)

- **Indeterminação parcial com referência explícita no contexto:** é possível encontrar no texto e/ou contexto qualquer elemento que nos forneça meios de inferência sobre o elemento gramatical que ocupa o lugar do referente.

Ex. (45)

INQ.- E no passado, o senhor acha que falava diferente assim...?

INF.- Eu acho que falava mais ainda diferente.

INQ.- É, né?

INF.- Depois dessa, dessa globalização, essa aldeia global que a gente vive. Eu acho que há uma tendência a unir mais esse sistema, né? Mas eu acho que antigamente aqui era, era difícil demais, era muito diferente. Sei lá, atualmente mesmo assim, se você for pra o interiô mermo, mermo com essa televisão jogando informações, quando a gente vai pra, pro interiô, a gente sente mermo que **O PESSOAL** do interiô fala aberto, né? (MA053/3)

A hipótese inicial que defendemos neste trabalho a respeito do *Grau de indeterminação do sujeito* é que há probabilidade encontrar maior número de ocorrências com a indeterminação parcial com referência explícita no contexto, visto que é possível encontrar no contexto qualquer elemento que auxilie na recuperação do referente. Assim, faz-se necessário reafirmar a importância deste estudo por meio dos fatores linguísticos e extralinguísticos selecionados.

Síntese do capítulo

O presente capítulo apresentou de forma detalhada as etapas necessárias para que esta pesquisa fosse desenvolvida, dando origem à dissertação. Aqui, especificamos a delimitação do perfil dos informantes, os instrumentos de pesquisa utilizados, a escolha das localidades, o *corpus* e tratamento quantitativo dos dados, dentre outras etapas. Após concluir esta etapa, demos continuidade ao estudo apresentando no capítulo a seguir a descrição e análise quantitativa e qualitativa dos dados sobre o uso das formas nominais no português falado nas capitais do Nordeste, considerando os fatores descritos neste capítulo.

4. EXPOSIÇÃO E ANÁLISE DOS DADOS

Após a organização dos dados, iniciamos a exposição dos resultados por meio de tabelas e gráficos, visando tornar mais dinâmico a sua interpretação. Para análise estatística, adotamos o programa computacional GoldVarbX (SANKOFF; TAGLIAMONTE; SMITH, 2005), que “mede os efeitos, bem como a significância dos efeitos, dessas variáveis independentes sobre a ocorrência das realizações da variável que está sendo tratada como dependente” (GUY; ZILLES, 2007, p. 105). O total geral foi de 392 ocorrências retiradas de 72 entrevistas do tipo DID, ou seja, diálogo entre informante e documentador.

Para fazer a análise dos dados, dividimos nossa variável dependente em dois grupos: Grupo A (artigo + substantivo) e o Grupo B (estrutura com um único elemento, ou seja, um substantivo), pois se referem às formas mais encontradas em nossos dados, como sinalizadas nos exemplos a seguir.

GRUPO A

Ex. (46)

INQ.– Que nomes dão a uma pessoa que bebe demais?

INF. – Beberrão

INQ. – E tem outro jeito...

INF. – **O CARA** bebe demais é um beberrão danado (AL077/3)

GRUPO B

Ex.(47)

INQ. – Chama de outro jeito

INF. – Chama. **NEGO** chama de sentina. Vaso sanitário. (BA93/3)

Em se tratando da organização dos dados, na primeira parte serão analisados os resultados referentes à distribuição total das ocorrências por grupo, indicado na subseção 4.1; na segunda, trataremos sobre os fatores selecionados pelo programa GoldVarb X, assinalados na subseção 4.2; e a terceira parte será destinada aos fatores que não foram selecionados pelo programa. Convém dizer que a explicação das variáveis em estudo obedecerá a ordem de escolha gerada pelo Programa.

4.1 Apresentação Geral dos Resultados

Em primeira análise, apresentamos os resultados gerais, destacando sua frequência de uso e as expomos, em ordem decrescente, conforme apresentada na Tabela 1.

Tabela 1 - Distribuição total das ocorrências por grupo

FNs		Nº ocorrências por FNs	Nº de ocorrências total	Percentual
GRUPO A	<i>O pessoal</i> <i>Os pessoal</i>	155	345	88%
	<i>A pessoa</i> <i>As pessoas</i> <i>As pessoa</i>	114		
	<i>O povo</i> <i>O povão</i>	49		
	<i>Os cara</i> <i>O cara</i>	12		
	<i>Os meninos</i> <i>Os menino</i>	6		
	<i>O rapaz</i> <i>Os rapaz</i>	6		
	<i>O jovem</i> <i>Os jovens</i>	3		
	<i>A colega</i> <i>Os colegas</i>	2		
	<i>Os moleque</i>	1		
	<i>Os pião</i>	1		
GRUPO B	<i>Fulana</i> <i>Fulano</i>	19	47	12%
	<i>Nego</i>	17		
	<i>Pessoas</i>	8		
	<i>Gente</i>	2		
	<i>Moleque</i>	1		
TOTAL			392	100%

Fonte: Elaborada pela autora.

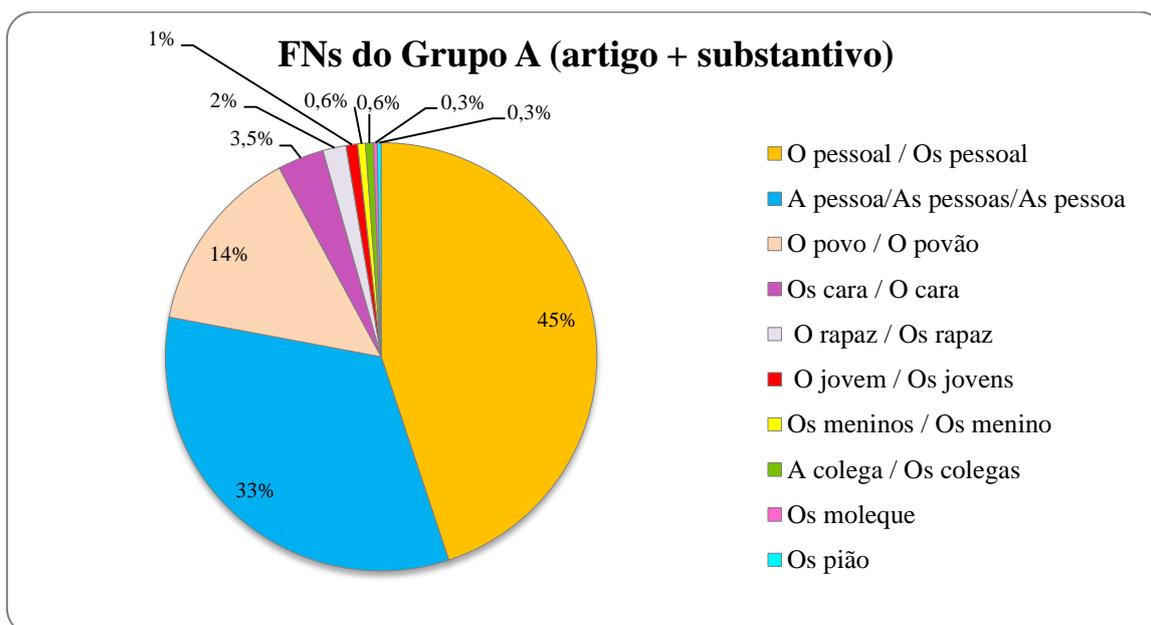
Conforme rodada feita pelo programa, constatamos que o Grupo A foi mais produtivo, com o registro de dez tipos diferentes de FNs, distribuídas entre 345 ocorrências, registrando percentual de 88%. Já o Grupo B apresentou cinco tipos diferentes de FNs, a partir de 47 ocorrências, representando 12% dos resultados. Os dados indicam uma distribuição significativa para o uso das FNs do Grupo A, ou seja, há uma maior recorrência de formas nominais quando seguidas de (artigo + substantivo).

Ainda de acordo com a Tabela 1, as FNs do Grupo A mais produtivas são respectivamente: i) *O pessoal/ Os pessoal*, com um total de 155 ocorrências; ii) *A pessoa/ As pessoas/ As pessoa*, computando 114 realizações, e iii) *O povo/ O povão*, com 49 ocorrências.

Em relação às variantes menos usadas pelos sujeitos desta pesquisa, se comparadas às formas mais produtivas encontradas no Grupo A, destacamos: *O cara/ Os cara, Os meninos/ Os menino, O rapaz/ Os rapaz, O jovem/ Os jovens, A colega/ Os colega, Os moleque* e *Os pião*. Convém dizer que, das sete menos produtivas neste grupo, as FNs *os pião* e *os moleque* foram as que tiveram o menor número de ocorrência, sendo registrado apenas um dado para cada forma.

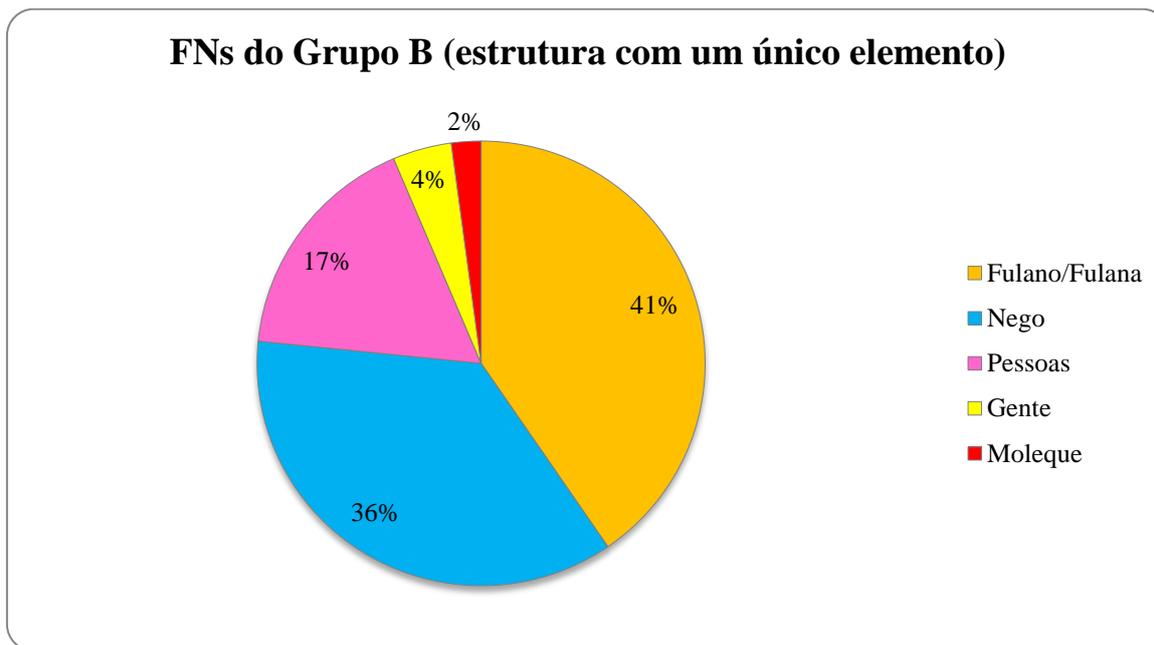
É possível ver no Gráfico 1, a seguir, o percentual que cada FN do Grupo A apresentou.

Gráfico 1 - Distribuição dos recursos de indeterminação da referência do sujeito do Grupo A



Fonte: Elaborado pela autora.

A respeito do Grupo B, as variantes mais usadas foram *Fulana/Fulano* - 19 ocorrências e *nego* com 17 realizações. Já em relação às menos produtivas, temos *Gente*, com duas ocorrências e *moleque*, com ocorrência única. Convém dizer que embora o pronome “a gente” também seja considerado uma forma de indeterminar a referência do sujeito no discurso, a variante que estamos considerando para no nosso estudo é a forma “*Gente*” por se tratar de uma FN. No Gráfico 2 notamos, em termos percentuais, a distribuição das FNs do Grupo B.

Gráfico 2 - Distribuição dos recursos de indeterminação da referência do sujeito do Grupo B

Fonte: Elaborada pela autora.

Para se entender melhor as FNs sentimos a necessidade de fazer uma busca mais profunda para compreender como essas palavras passaram a compor o universo em que a indeterminação do sujeito se encontra.

Tendo em vista que gramaticalização ocorre, em linhas gerais, quando um item lexical se torna gramatical, a depender das circunstâncias em que foi inserido²⁴ e de que a “lexicologia estuda o universo de todas as palavras, vistas em sua estruturação, funcionamento e mudança” (BARBOSA, 1990, p. 157), serão apresentadas ao longo de nossas análises o conceito das FNs que se mostraram mais expressivas, a fim de reforçar a ideia de que as palavras podem ganhar novos conceitos e assumir novas funções gramaticais, a depender do contexto em que estão inseridas, e que as gramáticas tradicionais, de modo geral, sentem dificuldade em considerar as novas abordagens de usos da língua.

Vejam agora algumas considerações sobre as FNs mais recorrentes que foram usadas como estratégias para indeterminar a referência do sujeito em nossa amostra:

O pessoal / Os pessoal – a palavra pessoal tem como remissão à palavra “pessoa” que vem do latim, “persona”, e corresponde ao papel que um ator desempenha, personagem

²⁴ Torna-se mais gramatical significa dizer que o item passa a assumir posições mais fixas nas sentenças, tornando-se previsível em termos de uso (MARTELOTTA *et al*, 1996, p. 46)

(CUNHA, 1996; FERREIRA, 1999; HOUAISS, 2001). Em outros dicionários, como Michaelis (1998), Borba (2004) e Aulete (2011), encontramos uma acepção que diz respeito a um conjunto de indivíduos que se reúnem por qualquer motivo. Convém dizer que as formas encontradas, neste estudo, não dizem respeito ao uso “pessoa” em si, mas sim, à forma empregada com o artigo definido, como “a pessoa”, “as pessoas” ou “o pessoal”. Quando usadas no singular, fica mais fácil identificar quem é o referente, no entanto, em construções como “as pessoas”, no plural, ou “o pessoal”, aqui no singular, mas com ideia de pluralidade, há generalização do referente, logo a ação verbal pode ser atribuída a qualquer pessoa, o que resultaria em um referente completamente genérico. Vejamos exemplos extraídos do nosso *corpus*:

Ex. (48)

INQ. – E a pessoa que deixa suas contas penduradas, aquela pessoa que não paga suas contas, como é que a gente chama também?

INF. – Esse é trambiqueiro, esse é... Virgem Maria! Esse é trambiqueiro, trapaceiro Eh... **O PESSOAL** chama carga torta. Uhn... Nó de... Nó cego. (MA026/8)

Ex. (49)

INQ. – Eu vou repetir a pergunta: aquela raiz que, coberta por uma casca marrom e que cozinha pra comer?

INF. – Macaxera ou ô aipim

INQ. – Aqui, chama das duas maneiras?

INF. – É

INQ. – Qual é que é mais comum a...

INF. – Macaxeira

INQ. – Aqui...

INF. – É

INQ. – Em Aracaju?

INF. – Uhn, uhn.

INQ. – Mas vocês falam também aipim?

INF. – É. A gente da, da época d’agora chamava macaxera e **O PESSOAL** mais antigo chama aipim. (SE079/2)

Em (48), vemos que a FN *o pessoal* possui um referente mais genérico que em (49), pois, ao ser questionado pelo inquiridor sobre a pessoa que tem o costume de pendurar as contas e não pagar, o informante deixa evidente que qualquer pessoa poderia dizer *carga torta*, não especificando a referência, diferente do exemplo (48), em que a FN *o pessoal* vem acompanhada de um adjunto adnominal, *antigo*, que o torna menos genérica. Além disso, é possível recuperar no contexto o elemento localizador *aqui* que limita mais ainda a referência desse sujeito, logo podemos ver que *o pessoal*

mais antigo pode ser ou não de Aracaju, uma vez que é possível limitar no espaço e tempo a referência desse sujeito.

Após a exposição geral do resultado por Grupos, partimos para uma outra apresentação dos resultados: procuramos, agora, observar como os Grupos de FNs se comportam em relação aos fatores linguísticos e extralinguísticos, visando chegar a uma análise binária. Para tanto, consideramos como grupo de aplicação as formas seguidas por determinantes, ou seja, o Grupo A, uma vez que as formas desse grupo apresentam maior frequência de uso e, conseqüentemente, o maior número de variantes.

Para uma análise mais refinada, foi preciso fazer várias rodadas testes para se chegar à rodada mais precisa na qual pudéssemos obter o peso relativo (PR)²⁵, fizemos primeiro uma rodada geral para verificar a frequência de uso entre os dois Grupos. Com a ocorrência de *KnockOut* no fator *Localidade*, tivemos que excluir a capital João Pessoa. Além disso, excluimos também do fator *Tempo Verbal*, as ocorrências em os tempos verbais Gerúndio, Imperativo e Particípio. Os resultados que seguem consideram a nova rodada binária dos dados.

4.2 Variáveis selecionadas pelo programa

Tendo o Grupo A como o valor de aplicação, o programa selecionou como fatores que favorecem a indeterminação da referência do sujeito os seguintes grupos: localidade e sexo, configurando assim como um fenômeno motivado por fatores extralinguísticos. Nas seções seguintes, apresentamos separadamente a análise de cada grupo de fator, considerando a ordem de escolha dada pelo programa.

4.2.1 Localidade

O fator localidade foi o primeiro fator a ser selecionado pelo programa. Na Tabela 2, a seguir, estão distribuídos os resultados encontrados nas nove capitais investigadas.

Tabela 2 - Distribuição das ocorrências por localidade

LOCALIDADE	GRUPO A	Peso Relativo
	Nº de ocorrências	
Recife	91	0.79
Natal	28	0.61

²⁵ O peso de um fator é calculado pelo GoldVarb X, com base em um conjunto de dados, que indica o efeito desse fator sobre o uso da variante investigada nesse conjunto.

Maceió	41	0.53
Aracaju	31	0.37
São Luís	30	0.35
Salvador	74	0.33
Fortaleza	11	0.21
Teresina	12	0.18
João Pessoa	27	-
Input: 0.90		Significância:0,009

Fonte: Elaborado pela autora.

Como já sinalizado, para as nossas rodadas no GoldVarbX resolvemos considerar o Grupo A como regra de aplicação por esse ter o maior número de variantes. Sendo assim, no quantitativo geral dos dados da Tabela 2, constatamos, após rodada binária, uma significativa diferença entre as nove localidades: Recife com PR 0.79, Natal, com PR 0.61, e Maceió, com PR 0.53 que se destacam pela tendência de uso das FNs do Grupo A. Contudo, ao verificarmos os resultados de maneira isolada, ou seja, a distribuição das variantes em cada uma das localidades observamos que o PR de Teresina foi de 0.18, um valor relativamente baixo se comparado com os de Recife, Natal e Maceió.

Notamos que das localidades investigadas, apenas João Pessoa não obteve PR, pois tivemos que excluir da rodada, visto que sua regra de aplicação foi de 100% para as formas do Grupo A, ocasionando, assim, *KnockOut*. Isso nos revela que os falantes pessoenses considerados para esta pesquisa têm maior preferência para as FNs que acompanham o determinante.

Assim, observamos que por estarem diatopicamente mais próximas, as capitais Natal, João Pessoa, Recife e Maceió apresentam comportamentos linguísticos parecidos em relação ao uso das FNs do Grupo A, enquanto São Luís e Salvador, que estão situados geograficamente nos extremos da região nordestina, considerando a dialetal proposta por Nascentes (1922), demonstraram ter comportamento diversificado. Mediante o exposto, constatamos que a hipótese inicial defendida neste trabalho a respeito do fator localidade se confirma, visto que foi possível verificar em São Luís o registro de 11 diferentes tipos de FNs, o mesmo aconteceu com Salvador, com total de oito formas.

Para melhor visualizar esse comportamento linguístico, trouxemos o detalhamento na Tabela 3 dos resultados obtidos em cada variante pertencente ao Grupo A.

Tabela 3 - Distribuição das FNs do Grupo A por localidade

Localidade	Recife	Maceió	Salvador	Natal	São Luís	Aracaju	João Pessoa	Fortaleza	Teresina
GRUPO A	Nº %	Nº %	Nº %	Nº %	Nº %	Nº %	Nº %	Nº %	Nº %
<i>O pessoal</i>	11	21	17	13	11	10	5	4	2
<i>Os pessoal</i>	47%	13%	11%	8%	7%	7%	3%	3%	1%
<i>A pessoa</i>	12	11	39	10	11	2	15	7	7
<i>As pessoas</i>	11%	9%	34%	9%	10%	2%	13%	6%	6%
<i>As pessoa</i>									
<i>O povo</i>	3	8	9	4	2	17	5	-	1
<i>O povão</i>	6%	16%	18%	8%	4%	35%	10%	-	2%
<i>O rapaz</i>	2	-	-	-	1	1	2	-	-
<i>Os rapaz</i>	33%	-	-	-	17%	16%	33%	-	-
<i>Os meninos</i>	-	-	-	-	-	-	-	-	2
<i>Os menino</i>									100%
<i>Os cara</i>	2	1	7	-	1	1	-	-	-
<i>O cara</i>	17%	8%	58%	-	8%	8%	-	-	-
<i>O jovem</i>	-	-	-	1	2	-	-	-	-
<i>Os jovens</i>				33%	67%				
<i>A colega</i>	-	-	-	-	2	-	-	-	-
<i>Os colegas</i>					100%				
<i>Os moleque</i>	-	-	1 100%	-	-	-	-	-	-
<i>Os pião</i>	-	-	1 100%	-	-	-	-	-	-
TOTAL	28	41	74	28	30	31	27	11	12

Fonte: Elaborada pela autora.

Em análise mais detalhada, levando em consideração as FNs encontradas neste estudo, constatamos que todas as nove capitais do Nordeste fazem uso das variantes *O*

peçoal / Os peçoal, sendo Recife com o maior número, totalizando 47% ocorrências, já Teresina foi a que teve menor percentual, computando 1% de ocorrências.

A respeito das variantes *A peçoal/As peçoas/ As peçoal*, todas as localidades aqui trabalhadas registraram ocorrências, no entanto, a que mais se destacou foi Salvador, com maior número de dados encontrados, totalizando 34%; já em Aracaju foram registradas apenas 2% de ocorrências.

Com exceção de Fortaleza, as demais capitais registraram a FN *O peço/ O peçoal*. Dentre as oito localidades, Aracaju se destaca por apresentar maior número de ocorrências para essa forma, 35%. Convém dizer que a FN *O peçoal* foi registrado apenas em Teresina, por um informante do sexo masculino pertencente a faixa etária II.

Sobre a variante *O cara / Os cara* tivemos registros em São Luís, Recife, Maceió, Aracajú e Salvador, sendo a última computando 58% ocorrências. Em se tratando da variante *O jovem / Os jovens* seu registro ocorreu apenas em duas localidades: São Luís, com 67% e Natal, 33%. Com relação à FN *Os pião*, computamos 100% de ocorrência em Salvador.

No tocante as FNs *A colega / As colegas* tivemos registro de 100% de ocorrências para os falantes de São Luís. Em Teresina obtivemos 100% registros da variante *Os meninos / Os menino*. Também foi possível verificar que em Salvador localidade foi a única que fez uso da variante *Os moleque* com 100% de ocorrência. Sobre as FNs *O rapaz / Os rapaz*, houve registro apenas em quatro localidades, sendo João Pessoa e Recife as capitais que obtiveram registro de 33% em cada. Já São Luís e Aracaju foi registrado 17% em ambas as localidades.

Embora João Pessoa não tenha contemplado resultados na Tabela 2 com valor em PR, vemos na Tabela 4 que os falantes dessa localidade têm maior preferência por quatro das 10 FNs pertencentes ao Grupo A: *O rapaz/Os rapaz*, computando 33%, *O peço/O peçoal*, com 10%, *A peçoal/As peçoas/As peçoal*, totalizando 13% e *O peçoal/Os peçoal*, com 3%. Para as demais FNs não houve registro de ocorrências.

Em relação às variantes do Grupo B, dispomos na Tabela 4, a seguir, a distribuição por localidade.

Tabela 4 - Distribuição das FNs do Grupo B por localidade

Localidade	Aracaju	Teresina	Fortaleza	Maceió	São Luís	Salvador	Recife	Natal	João Pessoa
GRUPO B	Nº %	Nº %	Nº %	Nº %	Nº %	Nº %	Nº %	Nº %	Nº %
<i>Fulana</i>	5	4	3	3	2	1	1	-	-
<i>Fulano</i>	26%	21%	16%	16%	11%	5%	5%	-	-
<i>Nego</i>	-	-	-	-	1 6%	15 88%	-	1 6%	-
<i>Pessoas</i>	-	1 12%	1 12%	1 12%	2 26%	-	2 26%	1 12%	-
<i>Gente</i>	-	-	3 60%	-	2 40%	-	-	-	-
<i>Moleque</i>	-	1 100%	-	-	-	-	-	-	-
TOTAL	5	6	7	4	7	16	3	2	-

Fonte: Elaborada pela autora.

A respeito das FNs que compõem o Grupo B, notamos que entre os falantes de Natal, Salvador e São Luís foi possível encontrar dados sobre a variante *Nego*, sendo Salvador a localidade que mais se destacou, totalizando 88% de ocorrências. Com exceção de Natal e João Pessoa, as demais capitais do Nordeste tiveram registro da variante *Fulano / Fulana*, sendo Aracaju com o maior número, 26% de ocorrências. Em se tratando da FN *Pessoas*, São Luís e Recife apresentaram o mesmo percentual, sendo registrado 25%, assim como Teresina, Fortaleza, Maceió e Natal que registraram 12% para cada localidade. Quanto a variante *Gente*, apenas duas capitais tiveram registro: Fortaleza com 60% e São Luís com 40%. Para *Moleque* tivemos 100% das ocorrências em Teresina.

4.2.2 Sexo

Em nossa amostra, essa variável ocupou o segundo lugar entre os fatores selecionados pelo programa. É possível ver na Tabela 5, a seguir, o panorama geral dos dados referente ao fator sexo.

Tabela 5 - Distribuição das FNs segundo sexo

SEXO	GRUPO A		Peso Relativo
	N ^a de ocorrências		
Mulher	171		0.59
Homem	174		0.40

Input: 0.90 **Significância:**0,009

Fonte: Elaborado pela autora.

Ao examinar a distribuição dos dados na Tabela 5 notamos que os sujeitos do sexo femininos tendem a fazer maior uso das FNs (0.59), enquanto os homens fazem menor uso dessas estratégias para indeterminar a referência do sujeito, com PR (0.40). Vemos que as mulheres têm uma maior tendência em marcar com o determinante as formas nominais de indeterminação da referência do sujeito. Fato este que encontra sua justificativa na ideia de que segundo o que prescreve a gramática normativa o determinante além de agir semanticamente sobre o nome que os segue, também precisa do contexto de uso, podendo ocupar a posição de determinante para especificar uma expressão nominal e despertar no ouvinte a consciência de determinado item já conhecido, como também pode servir com elemento genérico no sintagma nominal.

Na Tabela 6, a seguir, temos a interpretação dos dados mais detalhadas, agora considerando os números de ocorrências e percentual, de acordo com a distribuição total das FNs segundo o fator o sexo.

Tabela 6 - Distribuição das FNs segundo o fator sexo

GRUPOS	FNs	MULHERES		HOMENS		TOTAL
		N ^o	%	N ^o	%	
GRUPO A	<i>O pessoal</i> <i>Os pessoal</i>	57	33%	98	56%	155
	<i>A pessoa</i> <i>As pessoas</i> <i>As pessoa</i>	58	34%	56	32%	114
	<i>O povo</i> <i>O povão</i>	45	26%	4	2%	49
	<i>O rapaz</i> <i>Os rapaz</i>	5	3%	1	1%	5
	<i>Os meninos</i> <i>Os menino</i>	2	2%	-	-	2
	<i>A colega</i> <i>Os colegas</i>	2	2%	-	-	2

	<i>Os moleque</i>	1	1%	-	-	12
	<i>Os cara O cara</i>	1	1%	11	6%	4
	<i>O jovem Os jovens</i>	-	-	3	2%	1
	<i>Os pião</i>	-	-	1	1%	1
GRUPO B	<i>Fulana Fulano</i>	8	48%	11	46%	19
	<i>Nego</i>	8	48%	9	38%	17
	<i>Pessoas</i>	1	6%	1	4%	2
	<i>Gente</i>	-	-	2	8%	2
	<i>Moleque</i>	-	-	1	4%	1
	TOTAL	188	100%	204	100%	394

Fonte: Elaborada pela autora.

Ao examinar a Tabela 6, constatamos que as variantes pertencentes ao Grupo A mais produtivas para ambos os sexos foram: *O pessoal/Os pessoal*, com 155 e *A pessoa/As pessoas/ As pessoa*, com 114 ocorrências. Na fala dos homens desta pesquisa, observamos que as FNs *O pessoal/ Os pessoal* alcançou o percentual de 56% de ocorrências, já para as mulheres foi de 33%. Em se tratando das FNs *A pessoa/As pessoas/ As pessoa*, obtivemos maior resultado para as mulheres, com 34% de ocorrências e os homens com 32%.

Quanto a variante *O povo/O povão*, encontramos maior registro na fala de sujeitos do sexo feminino, com 26%. Em relação aos homens, embora tenhamos encontrado apenas 2% dos registros, foi em um informante do sexo masculino, segunda faixa etária e nível de escolaridade universitária, que foi coletado o uso da variante *o povão*. Na fala das mulheres, as FNs *Os meninos/ Os menino* e *A colega/As colegas* tiveram 2% de ocorrência por forma, enquanto *Os moleque* e *Os cara/ O cara* foram registradas apenas 1% forma para cada variante, enquanto *O jovem/Os jovens* não foram documentadas.

Em relação à fala dos homens, vemos que a variante *Os cara/ O cara* registraram 6% de ocorrências, já para *Os jovens/O jovem* encontramos 2% de ocorrências, enquanto para *O rapaz/Os rapaz* e *Os pião* somente 1% registro cada. Para

as variantes *Os meninos/Os menino*, *A colega/As colegas* e *Os moleque* não tivemos registros de ocorrências.

Em relação as FNs do Grupo B, notamos que as mais recorrentes entre as mulheres são *Fulana/Fulano* e *Nego*, computando 48% das ocorrências. O mesmo aconteceu para os homens, que também tiveram essas duas variantes como mais preferidas entre eles, sendo registrado 46% e 38% de ocorrências, respectivamente. Para as variantes *Pessoa*, *Gente* e *Moleque*, não foram documentadas na fala das mulheres, no entanto para os homens tivemos registro de 8% para *Pessoa* e 4% para *Gente* e *Moleque*, respectivamente.

A respeito da FN *O povão*, fizemos uma busca nos dicionários para entender a etimologia e o significado da palavra. Encontramos registro em três dicionários, Houaiss (2001), Borba (2004) e Aulete (2011):

O povão – a palavra “povão” aqui representada pelo acréscimo do sufixo -ão, advém da palavra “*povo +ão*”, e refere-se a “classe humilde, de baixa renda em oposição às classes médias e altas.” (AULETE, 2011, p.1095). Em outra obra encontramos a seguinte definição “grande quantidade de pessoas” (HOUAISS, 2001, p.2274) ou também, “multidão” (BORBA, 2004, p.1104). Observando nossos dados, vemos que se aproximam do sentido apresentado por Aulete (2011), em referência a um grupo de pessoas de classe mais humilde, conforme podemos ver no exemplo a seguir.

Ex. (50)

INQ. – E as festas que é que acontecia nas festas?

INF. – As festas. Nós tínhamos um jóquei clube, o jóquei, clube do jóquei que era elitista como ainda hoje é, e o clube dos diários que vinha dos diários associados que é onde **O POVÃO** podia entrar, com a sede... aqui próximo ((inint)) lá do outro lado do rio que era o balneário, praticamente, e o clube no centro da cidade, frequentava o comerciário todo mundo participava desse clube, e o jóquei clube, em termos de festa e as novenas na igreja do...do...das dores que é quem comandava o mês de setembro, esqueci até de falar: setembro, o mês das dores, as novenas comandavam... (PI034/7)

Em (50) é possível ver que a FN *O povão*, ainda que utilizada como forma indefinida, apresenta um referente mais específico, ou seja, ao ser questionado pelo inquiridor sobre o que costumava acontecer nas festas, o informante deixa evidente que havia dois tipos de clube, um mais elitista, frequentado apenas pelas pessoas abastardas, e o outro frequentado por associados com poder aquisitivo mais baixo, ou seja, *O povão*.

4.3 Grupos de fatores não selecionados

Dos grupos de fatores que considerávamos serem possíveis condicionadores do uso das FNs, tal como apontado em alguns estudos (SANTANA, 2006, ASSUNÇÃO, 2012; GUIMARÃES, 2018), quatro não foram selecionados em nosso estudo. São eles: *faixa etária, escolaridade, tempo verbal e grau de indeterminação*. Nas seções, a seguir, apresentaremos comentários sobre cada fator que, embora não selecionados pelo programa, merecem uma atenção sobre os resultados encontrados.

4.3.1 Faixa etária

O primeiro grupo a qual terçemos comentários é a faixa etária. Na Tabela 7, a seguir, temos a distribuição dos dados encontrados.

Tabela 7 - Distribuição das FNs segundo faixa etária

FAIXA ETÁRIA	GRUPO A	Peso Relativo	GRUPO B	Peso Relativo
	N ^a de ocorrências		N ^a de ocorrências	
Faixa I (18 a 30 anos)	126	[0.62]* ²⁶	10	[0.38]
Faixa II (50 a 65 anos)	219	[0.43]	37	[0.57]
Input: 0.88			Significância:0,036	

Fonte: Elaborado pela autora.

Ao examinar a Tabela 7, os resultados gerados pelo programa apontam, de modo geral, uma diferença não muito significativa entre as duas faixas etárias. No entanto, ao considerar os PR's, notamos que os sujeitos da faixa I são maiores usuários das FNs, com PR 0.62 do que os da faixa II, com PR 0.43. Em consideração ao exposto, Naro (2003, p. 40) afirma que

“Os falantes adultos tendem a preferir as formas antigas, criando uma situação estranha, pelo menos à primeira vista: existem pessoas que, apesar de estarem em interação constante (do tipo pai/filho), costumam falar de maneira distinta. Entretanto, isso não chega a comprometer a comunicação, já que ambos os lados são capazes de utilizar e entender todas as formas. Trata-se apenas de uma tendência em direção a outra forma. Com o correr do tempo, é provável que a forma nova seja adotada por todos.”

²⁶ Os pesos relativos entre colchetes se referem aos grupos não selecionados pela rodada binária do programa.

A seguir, na Tabela 8, apresentamos os dados interpretados de forma mais detalhada, agora considerando os números de ocorrências e percentual, de acordo com a distribuição total das FNs segundo o fator faixa etária.

Tabela 8 - Distribuição das FNs segundo a faixa etária

GRUPOS	FNs	FAIXA I		FAIXA II	
		Nº	%	Nº	%
GRUPO A	<i>A pessoa</i> <i>As pessoas</i> <i>As pessoa</i>	53	42%	61	28%
	<i>O pessoal</i> <i>Os pessoal</i>	39	31%	116	52%
	<i>O povo</i> <i>O povão</i>	24	119%	25	11%
	<i>Os cara</i> <i>O cara</i>	3	2%	9	4%
	<i>O rapaz</i> <i>Os rapaz</i>	3	2%	3	2%
	<i>O jovem</i> <i>Os jovens</i>	2	1%	0	0%
	<i>Os moleque</i>	1	1%	1	1%
	<i>A colega</i> <i>Os colegas</i>	1	1%	3	2%
	<i>Os meninos</i> <i>Os menino</i>	-	-	2	1%
	<i>Os pião</i>	-	-	1	1%
GRUPO B	<i>Pessoas</i>	5	50%	3	8%
	<i>Fulana</i> <i>Fulano</i>	3	30%	16	43%
	<i>Gente</i>	2	20%	-	0%
	<i>Nego</i>	-	-	17	46%
	<i>Moleque</i>	-	-	1	3%
	TOTAL	136	100%	256	100%

Fonte: Elaborado pela autora.

Ao examinar a Tabela 8, notamos que no Grupo A as duas FNs mais produtiva para a faixa I foram *A pessoa/ As pessoas/ As pessoa* e *O pessoal/ Os pessoal*, totalizando 42% e 31% de ocorrências, respectivamente, enquanto para a faixa II

tivemos mais registros da variante *O pessoal/ Os pessoal e A pessoa/ As pessoas/ As pessoa*, com 53% e 28%, respectivamente.

Sobre a forma *Nego* e *Os pião*, vemos que essas variantes se mostraram característica de fala de um informante da segunda faixa etária, com ensino fundamental incompleto, pertencente à cidade de Salvador. Nesse sentido, julgamos pertinente trazer algumas considerações de modo que fosse possível conhecer um pouco mais sobre a etimologia da palavra e seu significado.

Nego – a palavra *nego* tem sua origem em negro: negro > nego. Convém assinalar que a perda do /r/, supostamente, pode ocorrer pelo fato do falante achar que o uso da palavra sem o fonema pode ter uma conotação menos agressiva e preconceituosa. A palavra tem origem latina (*niger, nira, nigrum*) e diz respeito ao “indivíduo de etnia negra” (CUNHA, 1996). Apesar de fazer referência à etnia²⁷ negra, seu conteúdo semântico ampliou, assumindo caráter emocional em Salvador, por exemplo, podendo ser empregada para pessoas mais íntimas às quais se tenha estima ou para um referente mais genérico, ou seja, indeterminado, como no exemplo a seguir.

Ex. (51)

INF. – Cabriola. (inint) chama cabriola

INQ. – Uhn.

INF. – Sai rodano assim.

INQ. – É.

INF. – Chega e cai sentado.

INQ. – Isso.

INF. – **NEGO** chama cabriola (BA93/3)

Em relação à FN *os pião*, temos as seguintes informações:

Os Pião(s) – a palavra “pião”, aqui representada pelo alteamento da vogal pretônica, advém da palavra “peão”, e denomina o “indivíduo recrutado, em geral em outro Estado, como mão-de-obra para grandes empresas” ou, também, “servente de obra” (FERREIRA, 1999, p. 326). Convém dizer que o *peão* não possui uma profissão definida, e que isso se justifica pelas peculiaridades de seu trabalho. Em nossos dados, foi mantido o sentido original da palavra, no entanto, ela foi usada pelo informante 3

²⁷ Considerando que o termo “etnia negra”, trazido por Cunha (1996), seja reflexo de um texto escrito há aproximadamente 25 anos, optamos por não associar a lexia etnia à negra, visto que o termo não é sinônimo de raça, pois a palavra raça tem um sentido exclusivamente biológico, enquanto o termo etnia está cunhado para dar conta da diversidade cultural humana.

como recurso para indeterminar a referência do sujeito. Podemos constatar com um exemplo extraído do nosso *corpus*.

Ex. (52)

INQ. – São duas horas, né?

INF. – É, rapidium.

INQ. – Duas horas.

INF. – Então, pur'essa razão ainda não passiei de avião.

INQ. – É.

INF. – Mas lá **OS PIÃO** vai constantemente viajá de avião. (BA93/3)

Assim, vemos que os exemplos (51) e (52) nos mostram que o contexto a qual as FNs estão inseridas influencia na escolha e no uso das estratégias para indeterminar a referência do sujeito.

Em relação as variantes *O povo/ O povão*, não tivemos muita diferença entre as duas faixas etárias, entretanto, a faixa I teve maior número de ocorrências, totalizando 19%, já para a faixa II foi computado 11%.

Ainda de acordo com a Tabela 8, as FNs menos usadas pelos sujeitos da faixa etária I foram *O rapaz/ Os rapaz* e *Os cara/ O cara* totalizando 2% para cada FNs. Convém dizer que das cinco formas menos produtivas, destacamos três FNs que tiveram apenas 1% de ocorrência: *As colega/ Os colega* e *Os moleque*. Para essa faixa, não tivemos registro de *Os meninos/ Os menino* e *Os pião*.

Sobre a faixa II, temos como menos produtivas as seguintes variantes, respectivamente: *Os rapaz/ O rapaz* e *A colega/ As colega* com 2% de ocorrências para cada FN. Já para as FNs *Os meninos/ Os menino*, *Os pião* e *Os meninos* tivemos 1% das ocorrências. Sobre a FN *Os cara/ O cara* tivemos registro de 4% de ocorrências, ou seja, seis ocorrências a mais em relação à faixa I.

Ainda em relação a faixa etária II, ao examinar as FNs do Grupo B, notamos que *Pessoas* e *Moleque* não se mostraram muito produtivas, sendo encontradas apenas 8% e 3%, respectivamente. Sobre as FNs *Gente* não encontramos registro. Constatamos ainda que as FNs mais produtivas no Grupo B foram: *Nego* e *Fulano/Fulana*, sendo computado 46% e 43% de ocorrências, respectivamente.

A respeito da faixa I, notamos que a FNs que se mostrou mais produtiva foi *Pessoas*, com registro de 50%, seguido de *Fulano / Fulana*, com 30% e *Gente* totalizando 20%. Para *Nego* e *Moleque* não houve ocorrência.

Visando obter resultados mais precisos, optamos por fazer o cruzamento da variável sexo em relação à faixa etária, com o objetivo de verificar se os dois fatores em questão, quando combinados, fornecem informações mais precisas sobre as formas nominais de indeterminação da referência do sujeito. Ao fazer o cruzamento entre as duas variáveis, obtemos os seguintes resultados representados na tabela abaixo.

Tabela 9 - Cruzamento entre sexo e faixa etária

SEXO	MULHER		HOMEM	
FAIXA ETÁRIA	GRUPO A	GRUPO B	GRUPO A	GRUPO B
Faixa I	95%	5%	90%	10%
Faixa II	88%	12%	84%	16%

Fonte: Elaborado pela autora.

Após observar os números e analisar os resultados, notamos que, proporcionalmente, tanto homens quanto mulheres, fazem maior uso das FNs do Grupo A. Observamos também que não houve um resultado significativo em relação às faixas etárias, no entanto vemos que as mulheres jovens (18 a 35 anos) usam com maior frequência as variantes seguidas de determinantes, ou seja, FNs como *O pessoal e A pessoa*, por exemplo. Os resultados corroboram com os selecionados pelo programa.

4.3.2 Escolaridade

Aos olhos dos linguistas brasileiros (PAIVA; SCHERRE, 1999; BORTONIRICARDO, 2004) parece ser de comum acordo a ideia de que o grau de escolaridade pode ser visto como indicador para ajudar a esclarecer as diferenças entre o português padrão e o português popular. Segundo Oliveira e Silva & Paiva (1996, p. 343), “em países em que a escolarização não é apanágio da(s) classe(s) mais favorecidas(s), é a variável classe social que tem preferentemente servido de marcador”.

Assim como o fator faixa etária, a variável escolaridade não foi selecionada pelo programa durante as rodadas, todavia é conveniente que façamos a interpretação desses resultados para entender como as FNs são entendidas considerando o grau de aceitabilidade das FNs nas localidades trabalhadas nesta pesquisa, tendo em vista que o fenômeno em estudo trata-se de variantes inovadoras usadas para indeterminar a referência do sujeito em contexto de fala/escrita. O objetivo aqui é, primeiramente, investigar o efeito isolado dessa variável para, em seguida, examinar o cruzamento dela

com a variável sexo. Para tanto, apresentamos na Tabela 10 os resultados obtidos por nossa pesquisa.

Tabela 10 - Distribuição das FNs segundo a escolaridade

ESCOLARIDADE	GRUPO A	Peso Relativo	GRUPO B	Peso Relativo
	N ^a de ocorrências		N ^a de ocorrências	
Fundamental	194	[0.50]	26	[0.50]
Universitário	153	[0.49]	21	[0.51]
<i>Input: 0.88</i>			<i>Significância: 0.969</i>	

Fonte: Elaborado pela autora.

Como podemos observar na Tabela 10, embora não haja diferença muito significativa, os falantes com nível fundamental tendem a fazer maior uso das FNs, sendo registrado PR 0.50; já para os do nível universitário: PR 0.49. Verificamos que os resultados obtidos não corroboram com aquilo defendido em nossa hipótese de que quanto maior a escolarização, mais favorável será o uso de recursos para indeterminar a referência do sujeito. Sobre isso, Bortoni-Ricardo (2004, p. 48) salienta que “os anos de escolarização de um indivíduo e a qualidade das escolas que frequentou também têm influência em seu repertório sociolinguístico. Observa-se que esses fatores estão intimamente ligados ao estatuto socioeconômico, na sociedade brasileira”.

A seguir, na Tabela 11, expomos os dados, agora de forma mais detalhada, considerando os números de ocorrências e percentuais presentes nos dois grupos de FNs.

Tabela 11 - Distribuição das FNs segundo a escolaridade, em percentual

GRUPOS	FNs	FUNDAMENTAL		UNIVERSITÁRIO	
		Nº	%	Nº	%
GRUPO A	<i>O pessoal Os pessoal</i>	104	54%	51	33%
	<i>A pessoa As pessoas As pessoa</i>	45	23%	69	45%
	<i>O povo O povão</i>	23	12%	26	17%
	<i>Os cara O cara</i>	9	4%	3	2%
	<i>O rapaz Os rapaz</i>	6	3%	-	-

	<i>Os meninos Os menino</i>	2	1%	-	-
	<i>A colega Os colegas</i>	1	1%	1	1%
	<i>O jovem Os jovens</i>	1	1%	2	2%
	<i>Os pião</i>	1	1%	-	-
	<i>Os moleque</i>	-	-	1	1%
GRUPO B	<i>Fulana Fulano</i>	10	39%	9	43%
	<i>Nego</i>	8	31%	9	43%
	<i>Pessoas</i>	6	23%	2	10%
	<i>Gente</i>	2	7%	1	4%
	<i>Moleque</i>	-	-	-	-
	TOTAL	218	100%	174	100%

Fonte: Elaborado pela autora.

Ao fazer a leitura da Tabela 11, constatamos que as FNs pertencentes ao Grupo A mais produtivas para os sujeitos com ensino fundamental incompleto foram respectivamente: *O pessoa/Os pessoal*, com total de 54%, *A pessoa/ As pessoa/ As pessoa* com 23% e *O povo/O povão* com 12%. Em relação às FNs menos produtivas, tivemos na seguinte ordem: *Os caras/O cara*, totalizando 4%, *O rapaz/Os rapaz* com 3% das ocorrências e *Os meninos/Os menino*, *O jovem/Os jovens*, *A colega/As colegas* e *Os pião* computando 1% para casa FN. Não houve citações para *O Moleque* neste grupo de informantes.

Em relação ao nível universitário, as FNs mais recorrente são: *A pessoa/ As pessoa/ As pessoa* com 45%, *O pessoal/Os pessoal*, com total de 33% de ocorrências e *O povo/O povão*, computando 17%. Sobre as FNs com baixo número de ocorrências, respectivamente tivemos: *O rapaz/ Os rapazo* totalizando 3% e *Os cara/ O cara* com 2% das realizações. Para *A colega/ As colega*, *Os menino/Os menino* e *O jovem/ Os jovens* foi computado apenas 1%. *Os pião* não foram registradas neste recorte de escolaridade.

Ainda sobre a Tabela 11, agora analisando dos dados do Grupo B, notamos que as FNs mais recorrente na fala dos sujeitos com nível fundamental foram,

Fulano/Fulana, com 39%, *Nego*, com 31% das realizações e *Pessoas* totalizando 23%. Já a FN *Gente* computou 7%. *Moleque* não houve registro na fala de pessoas com esse nível de escolaridade.

No que diz respeito ao nível universitário, as FNs *Fulano/Fulana* e *Nego* tiveram o mesmo número de ocorrências, computando 43% registros cada; já para a FNs *Pessoas* tivemos 10% das realizações. *Gente* se mostrou ser menos preferidas entre os sujeitos desta pesquisa, sendo registrado 4%. Convém dizer que não tivemos ocorrências para as FNs *Moleque*.

Em relação à FN *Fulano/Fulana*, mais recorrente entre os falantes dos dois níveis de escolaridade a qual citamos no título deste trabalho, trouxemos algumas informações que nos ajudam a entender o processo de indeterminação da palavra.

- ***Fulano/ Fulana***: provém do adjetivo árabe (fulân) e significa “tal”. Conforme a expansão árabe sobre a península Ibérica durante a Idade Média, o termo *fulân* foi sendo incorporado à língua castelhana e mantendo o mesmo significado. Os espanhóis, por volta do século XIII, usavam *fulano* como pronome, em construções sintáticas do tipo: *fulana casa (tal casa)*, *fulano sujeito (tal sujeito)*. No entanto, já se tinha registro do uso de *fulano* como adjetivo antes do século XV, com as obras de Picón Febres, escritor venezuelano, o novelista mexicano Dante Medida e no Dicionário Corominas-Pascual. Autores como Domínguez (1879 *apud* VELARDE, 2000, p. 4) afirma que os dicionários gerais da língua não trazem com clareza a definição de *fulano*, principalmente o dicionário acadêmico, aqui apresentado de forma parafraseada, como: voz com o qual o nome de uma pessoa é fornecido, quando é ignorado propositalmente sem deixar expresso. Em vez disso, o autor sugeriu a seguinte definição: “nome próprio de uma sugestão imaginária que é apresentada como um tipo em comparação, suposição, exemplos, etc. que são estabelecidas a fim de circunscrever ou especificar a ideia que é enunciada”²⁸. Convém dizer que na 5ª acepção trazida pelo *Diccionario de la lengua española*, a palavra *fulana* além de fazer referência a uma pessoa indeterminada, também pode ser usada como sinônimo para prostituta. No português, “fulano” virou substantivo, e até derivou para a forma “fuão” em Portugal. É possível encontrar em obras como

²⁸ ‘nombre propio de un sujeto imaginario que se presenta como tipo en las comparaciones, suposiciones, ejemplos, etc. que se establecen, a fin de circunscribir o concretar la idea que se enuncia’ (Diccionario, s. v. fulano).

de Machado (1990) acepções que afirmam haver interferência da língua castelhana na composição a palavra *fulano*, antes de ser documentada em português como *folam* e *fulam*. Aulete (2001, p.685) traz a seguinte definição para *fulano* “nome usado para se referir a uma pessoa desconhecida ou que não se quer nomear, um indivíduo qualquer que não se quer identificar”. Notamos que o significado se mantém ao se referir às FNs, uma vez que *fulano/fulana* é usado como estratégia para se indeterminar a referência do sujeito. A respeito disso, apresentamos a seguir um exemplo retirado do nosso *corpus*.

Ex. (53)

INF. – Corno

INQ. – Chama de outra maneira?

INF. – Não... Se diz qu'ê corno, fulano, **FULANO** levou gaia. Eu falo assim. (SE93/3)

Assim como nos outros fatores não selecionados pelo programa, também resolvemos fazer o cruzamento entre os dados para melhor entendimento dos resultados. Sendo assim, resolvemos cruzar o fator sexo com a escolaridade. Esses resultados estão dispostos na Tabela 12, a seguir.

Tabela 12 - Cruzamento entre sexo e nível de escolaridade

SEXO	MULHER		HOMEM	
	GRUPO A	GRUPO B	GRUPO A	GRUPO B
Fundamental	92%	8%	86%	14%
Universitário	90%	10%	85%	15%

Fonte: Elaborada pelo autora.

De acordo com a Tabela 12, notamos que os sujeitos do sexo feminino, são mais favoráveis ao uso das variantes do Grupo A. Isto é, na fala de mulheres com ensino fundamental incompleto tivemos o registro de 92%, seguido das mulheres com o ensino universitário, com 90% dos dados para FNs com determinantes. Observamos também que para ambos os níveis de escolaridade, os resultados encontrados na fala dos homens não apresentaram muita diferença, para o fundamental tivemos 86%, já o universitário 85% para o Grupo A. Notamos que de maneira geral, as FNs do Grupo B não apresentaram percentual muito elevado, deixando evidente uma média de 11,75% para esse tipo de realização.

4.3.4 Graus de indeterminação

Como já citado, o fator linguístico grau de indeterminação também não foi selecionado pelo programa, entretanto é pertinente fazermos algumas observações sobre essa variável tendo em vista que este se configura com um fator recorrente no que concerne os estudos sobre a indeterminação do sujeito. A seguir, na Tabela 13, apresentamos os resultados encontrados após rodada estatística.

Tabela 13 - Distribuição das FNs segundo Grau de Indeterminação

Grau	GRUPO A	Peso Relativo	GRUPO B	Peso Relativo
	Nº de ocorrências		Nº de ocorrências	
Indeterminação completa	81	[0.47]	12	[0.53]
Indeterminação parcial com referência implícita no contexto	111	[0.59]	10	[0.41]
Indeterminação parcial com referência explícita no contexto	153	[0.44]	25	[0.66]
<i>Input:0.88</i>			<i>Significância:0.287</i>	

Fonte: Elaborado pela autora

Embora nossa hipótese inicial seja que construções com indeterminação parcial com referência explícita no contexto favoreça o uso das FNs, notamos que os resultados dizem o contrário. Constatamos por meio dos PR's que os informantes desta pesquisa tendem a fazer maior uso da indeterminação parcial com referente implícito (0.59), seguido da indeterminação completa com PR (0.47) e em último a indeterminação parcial com referência explícita (0.44). Contudo, como os dados não permitam afirmar com segurança sobre a escolha do tipo de indeterminação, podemos sugerir que talvez, o informante não veja necessidade de utilizar algum tipo de referente durante o diálogo.

Assim, para melhor ilustrar, resolvemos apresentar a seguir exemplos extraídos do nosso *corpus*. Convém dizer que embora não deixe claro por se tratar de um recorte, nota-se discrepâncias entre os graus de indeterminação. Nos nossos dados, essa abrangência pode apresentar sentido mais amplo ou mais restrito, a depender do seu referente. Vejamos os exemplos (54), (55) e (56).

Ex. (54) - Indeterminação completa

INQ.– ... um rio pequeno, de uns dois metros de largura?

INF.– **GENTE** chama de riacho.

INQ.– Tem outro nome assim que você conheça?

INF.– Não, sempre só riacho mesmo e rios.

INQ.– Certo! (MA026/1)

Ex. (55) - Indeterminação parcial com referência implícita no contexto

INQ. – O senhor conhece alguma simpatia pra... por exemplo pra tirar verruga ou pra outra coisa qualquer?

INF. – Eh... é simpatia pra tirar berruga, né?

INQ. – Ou outra que o senhor conheça.

INF. – Outra coisa. Eu pelo menos tinha uma berruga no meu... no meu dedo aqui do pé e **O RAPAZ** me ensinou uma vez eu tava... eu fazia era cortar assim com estilete e ele disse: “ah, cara quanto mais tu corta mais... aí cresce mais. Sabe o que que tu faz? Tu pega um... um... pedaço de... de... tomate, tu corta um pedacinho assim bota e amarra pra não caí, né?” Olha, poheu botei isso uma semana e acabou. Eu disse: rapaz, teu remédio é bom mesmo. Tomate. Eu também tinha muito esse negócio, tinha uns pano branco que aparecia em mim, aí o rapaz disse assim: “Olha, tu compra sabe o quê? Enxofre.” Você conhece enxofre? (MA026/7)

Ex. (56) - Indeterminação parcial com referência explícita no contexto

INQ.– Ah, também? Você tem alguma explicação pra essa... barriga branca?

INF.– Num sei, os... os... os... **OS COLEGAS** na faculdade chamava... falavam que era aquele marido que... que a mulhé manda nele e... tal e... e ele é corno, então chama barriga branca. (MA026/8)

Em (54), observamos que o diálogo se apresenta de forma precisa, em que ao ser questionado pelo inquiridor sobre como se chama um rio pequeno, o informante utiliza a FN *Gente* para indeterminar a referência do sujeito e dizer que se chama riacho. Notamos que no exemplo em questão, não é possível recuperar o referente a quem se fala, visto que o contexto não é amplo o suficiente para inferir a quem *Gente* possa estar relacionado, logo o grau de indeterminação usado foi a completa.

Em (55) que ao ser questionado pelo inquiridor sobre se ele conhecia alguma simpatia para tirar verruga ou outra coisa qualquer, o informante contextualiza a resposta e, em seguida, utiliza a variante *O rapaz* para dizer que ele aprendeu com alguém, ou seja, notamos que há uma contextualização para só assim se chegar a FN em questão. Observamos que o grau de indeterminação encontrada neste trecho se refere a implícita, visto que não temos algum elemento prévio que nos sinalize quem ou de que lugar é *O rapaz* em questão.

No trecho (56) observamos que o informante fez uso de uma FN com determinante para não revelar a identidade do sujeito. Notamos que neste exemplo fica mais explícito a referência do sujeito, uma vez que ao ser questionado pelo entrevistador sobre se existe alguma explicação para o uso do termo “barriga branca” para designar o homem que recebe ordem da mulher, o informante usa a FN *Os colegas*. Como recurso

para restringir ainda mais a referência é usado a construção sintática “na faculdade” para sinalizar que não se trata de qualquer grupo de colegas, mas sim aqueles que fazem faculdade com o informante, logo fica evidente que a indeterminação é parcial com referente explícito.

4.3.5 Tempo verbal

Outro grupo não selecionado pelo programa foi o fator *Tempo verbal*. Contudo, sentimos a necessidade verificar que influência essa variável tem sob o fenômeno da indeterminação do sujeito, tendo em vista que em outros estudos (SANTANA, 2006, ASSUNÇÃO, 2012; GUIMARÃES, 2018) esse fator foi significativo para as análises. Os resultados desse grupo constam na Tabela 14, a seguir.

Tabela 14 - Distribuição das FNs segundo tempo verbal

TEMPO VERBAL	GRUPO A	Peso	GRUPO B	Peso
	Nº de ocorrências	Relativo	Nº de ocorrências	Relativo
Presente	278	[0.48]	39	[0.52]
Pretérito	50	[0.55]	6	[0.45]
Futuro	7	[0.65]	1	[0.35]
Input: 0.87			Significância: 0.942	

Fonte: Elaborado pela autora.

Verificamos que no Grupo A o tempo verbal que apresenta maior número de ocorrência foi o presente, no entanto quando vamos verificar os pesos, notamos que o Futuro (0.65) mais favoreceu o uso das FNs, seguido de ocorrências no Pretérito (0.55) e Presente no (0.48). Nos exemplos (50), (251) e (52), podemos observar como as FNs se comportam conforme o tempo verbal.

Ex. (57):

INQ. – Mas brasileiros, nascidos aqui no Rio Grande do Norte. Algum grupo que você acha que é diferente, que fala diferente?

INF. – Olha, assim tal... assim também por exemplo, **O JOVEM** não tem a mesma, mesma linguagem do adulto. De repente tem gírias próprias da idade, né? também. Qu'eu considero que seja diferente. Não é tanto diferente, né? (RN053/7)

Ex. (58):

INQ. – E no caso de ter caído, a mesma situação, da, da mão de uma mulher idosa já e como um rapaz jovem vai chamar?

INF. – De uma senhora, no caso, é seria mais ou menos isso: “Ei tia” ou “Vó”, dependendo se ela tiver os cabelinho muito brancos pode chamar até de Vó. **O JOVEM** passou a usar muito esses termos né, tio, tia, vô, vó. (MA026/7)

Ex (59):

INF. – Mandioca ô aipim

INQ. – Como é isso? Mandioca ou aipim, tanto faz chamar de um jeito ou de outro?

INF. – Dizem que existe variações, né?

INQ. – Uhn.

INF. – Inclusive, assim, que uma serve pra comê, ôtra não serve e tal. Mas eu cunheço... eu geralmente chamo de aipim

INQ. – Uhn.

INF. – Quando eu vô comprá, eu chamo de aipim mais ô... eu ôço **AS PESSOAS** também chamarem de mandioca. (BA93/5)

Em (57), temos o uso da FN *O Jovem* em contexto no presente, ou seja, relacionada à característica atemporal o fato ocorrido. No exemplo (58), a FN *O Jovem* foi empregada no pretérito indicando um fato que aconteceu em um determinado momento no passado. Já em (59), vemos o uso de *As pessoas* empregada no futuro, dando ideia de eventualidade e possibilidade.

Para analisar ainda mais este fator, resolvemos cruzá-lo com o grau de indeterminação. Os resultados estão descritos na Tabela 15, a seguir.

Tabela 15 - Distribuição das FNs segundo Tempo verbal

GRAU	Indeterminação completa		Indeterminação parcial com referência implícita no contexto		Indeterminação parcial com referência explícita no contexto	
	GRUPO A	GRUPO B	GRUPO A	GRUPO B	GRUPO A	GRUPO B
Presente	89%	11%	90%	10%	86%	14%
Pretérito	79%	21%	96%	4%	92%	8%
Futuro	100%	0%	0%	0%	75%	25%

Fonte: Elaborado pela autora.

De maneira geral, os resultados inclinam-se com maior preferência para as variáveis do Grupo A. Notamos que para a Indeterminação completa, tivemos 100% do

número de ocorrências para realizações no Futuro, seguido de 96% para os dados no Pretérito com grau a indeterminação parcial com referência implícita e 92% Indeterminação parcial com referência explícita. Podemos observar também que embora as ocorrências no Futuro tenham se destacado por apresentar o maior PR, notamos na tabela de cruzamento o percentual elevado para as ocorrências no Presente principalmente no grau de *Indeterminação parcial com referência implícita no contexto*.

Síntese do capítulo

Neste capítulo apresentamos a discussão a cerca dos dados coletados a partir de 72 entrevistas realizadas pelo Projeto ALiB com falantes das capitais nordestinas. Assim, submetemos os dados a rodadas estatísticas no Programa GoldVarb X dando ênfase para o Grupo A (artigo + substantivo), grupo usado como regra de aplicação e em seguida expomos por meios de quadros, tabelas e gráficos os resultados encontrados. Além disso, fizemos o cruzamento entre fatores a fim de verificar como os fatores selecionados para este estudo se comportam mediante a análises conjuntas.

5. CONSIDERAÇÕES FINAIS

O objetivo desta pesquisa é investigar as formas nominais de referenciação do sujeito, a fim de verificar a situação de uso na fala de nordestinos, com base nos dados do Projeto ALiB (Atlas Linguístico do Brasil), examinando quais fatores linguísticos ou sociais condicionam o uso das FNs, e, conseqüentemente, quais favorecem o processo de variação linguística.

No primeiro capítulo, discorremos sobre a indeterminação do sujeito, tomando como base o que postula as gramáticas normativas (CUNHA; CINTRA, 1985; CEGALLA, 2008). Apresentamos também considerações defendidas em manuais de linguística por alguns teóricos como (PERINI, 1995; DUARTE, 2007), além de expor os resultados de estudos que também investigaram a indeterminação no PB.

O terceiro capítulo destinou-se a apresentação dos procedimentos metodológicos adotados na pesquisa. Dissertamos a respeito do Projeto ALiB, sobre a história de fundação das localidades investigadas, da construção do nosso corpora usando dados de fala de informantes das capitais nordestinas. Apresentamos também os fatores analisados, com conceituação, exemplificação e hipóteses.

Conforme a análise dos 392 dados foi possível constatar que os falantes das capitais nordestinas possuem um acervo significativo de FNs para indeterminar a referência do sujeito. Para melhor compreensão e interpretação dos resultados, resolvemos dividir em dois grupos: Grupo A, destinado as variantes antecedidas de determinante e Grupo B, composta por estrutura com um único elemento.

Inicialmente fizemos um estudo do percentual geral das ocorrências para identificar aquela(s) mais recorrente entre os falantes nordestinos. Com base nas observações, foi possível elencar alguns pontos que se destacaram na análise dos dados. São eles:

- As FNs que compõem o Grupo A tiveram maior número de ocorrências, sendo registrado 345 (88%) dados; já o Grupo B teve o cômputo de 47 (12%). Dentre as FNs mais usadas pelos sujeitos dessa pesquisa, podemos destacar de modo geral as seguintes: *O pessoal/Os pessoal, A pessoa/As pessoas/As pessoa, O povo/O povão, Nego e Fulano.*
- Em rodada binária tivemos dois grupos selecionados pelo programa: Localidade e Sexo. Sobre as localidades, vimos que Recife, Natal e Maceió foram as capitais com maiores PR, sendo registrado respectivamente 0.79, 0.61 e 0.53.

- Constatamos que as mulheres, embora sejam consideradas responsáveis por conservar a norma padrão, vimos em nossas análises que elas fazem maior uso das FNs, sendo computado PR 0.59, contra PR 0.40 para os homens. Notamos também que as variantes pertencentes ao Grupo A mais preferidas entre ambos os sexos foram: *O pessoal/Os pessoal, A pessoa/As pessoas/ As pessoa*, sendo a primeira mais recorrente na fala dos homens. Já a FN *O povo/O povão*, teve maior registro na fala de mulheres. Quanto as outras variantes pertencentes ao Grupo A não tiveram registros muito significativos. Sobre o Grupo B, as mais recorrentes foram *Fulano/Fulana e Nego* entre a fala de homens e mulheres.
- Os fatores não selecionados pelo programa foram: Faixa Etária, Escolaridade, Tempo verbal e Grau de indeterminação. No que tange à faixa etária dos informantes, constatamos que a faixa I fez maior uso das FNs, computando PR [0.64] das ocorrências, contra PR [0.43] realizações da faixa II. Sobre as variantes do Grupo A, as mais produtivas para a faixa I foram respectivamente: *A pessoa/ As pessoas/ As pessoa e O pessoal/ Os pessoal*, já para a faixa II, os resultados foram inversos. Quanto as variantes do Grupo B, notamos que a FN *Nego* se mostrou mais produtiva na faixa II, enquanto para a faixa I não houve registros. O mesmo aconteceu com *Fulano / Fulana*, no entanto na fala dos mais jovens houve poucos registros.
- Embora a escolaridade ou nível de escolarização seja um fator de estratificação social, foi possível observar em nossos resultados uma leve tendência do uso das FNs por sujeitos com menor grau de instrução, sendo registrado PR [0.50] para o nível fundamental e PR [0.49] para o universitário. Como variantes mais recorrentes para os sujeitos com ensino fundamental tivemos: *O pessoa/Os pessoal, A pessoa/ As pessoa/ As pessoa, Os caras/O cara e O povo/O povão*, já para o nível universitário, as FNs mais recorrente são: *A pessoa/ As pessoa/ As pessoa, O pessoal/Os pessoal, e O povo/O povão*, ambas pertencentes ao Grupo A. Em relação ao Grupo B, notamos que as mais recorrentes na fala dos sujeitos com nível fundamental e universitário foram *Fulano/Fulana e Nego*.
- O Tempo Verbal que mais favoreceu o uso das FNs do Grupo A obedeceu a seguinte ordem: Futuro com PR de [0.65], Pretérito, com [0.55] e Presente [0.48]. Refutando a nossa hipótese de que construções no presente do indicativo possibilitam o uso das FNs.

- Para o fator Grau de Indeterminação notamos que as ocorrências com a Indeterminação parcial com referente implícito tiveram maior PR [0.59], seguido da indeterminação completa com PR [0.47] e em último a indeterminação parcial com referência explícita [0.44]. Fato esse que refuta a nossa hipótese de que é mais comum haver indeterminação quando a referência está explícita.

Considerando as hipóteses levantadas ao longo deste trabalho, notamos que a maioria delas não se confirmaram, apenas a que se refere a São Luís e Salvador terem mais diferentes tipos de formas nominais de referenciação. Percebemos também que os falantes das capitais nordestinas utilizam as FNs para indeterminar a referência do sujeito, constatando, assim, uma variedade de recurso de indeterminação no português falado na Região do Nordeste. É importante ressaltar que os resultados aqui apresentados devem ser encarados como tendências, dada a configuração metodológica do nosso corpus. Portanto, podemos concluir que as variantes inovadoras encontradas neste trabalho nos revelam ser comum, na fala de nordestinos, esses recursos de indeterminação que a gramática normativa ainda não contemplou. Por fim, acreditamos que este estudo servirá de subsídio para pesquisas futuras, tanto para o português falado no Brasil quanto no Nordeste.

REFERÊNCIAS

- ALBUQUERQUE JÚNIOR, Durval Muniz de. *A invenção do Nordeste e outras artes*. Prefácio de Margareth Rago, São Paulo: Cortez, 2009.
- ALKMIM, Mônica Guieiro Ramalho de. As negativas sentenciais no dialeto mineiro: uma abordagem variacionista. (2001) 261f. Tese (Doutorado em Linguística) Programa de PósGraduação em Letras: Estudos Lingüísticos, Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte.
- ALMEIDA, Napoleão Mendes de. *Gramática metódica da língua portuguesa*. 29ª ed São Paulo: Saraiva, 2009.
- ARAUJO, Sérgio Murilo Santos de. A região semiárida do Nordeste do Brasil: Questões Ambientais e Possibilidades de uso Sustentável dos Recursos. Rios Eletrônica- *Revista Científica da FASETE*, ano 5 n. 5 janeiro de 2011, p.89-98.
- ASSUNÇÃO, Janivam da Silva. *A indeterminação do sujeito na variedade linguística de Feira de Santana: um estudo variacionista*. 2012. 117 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual de Feira de Santana, Feira de Santana, 2012.
- Atlas Linguístico-Etnográfico de Portugal e da Galiza (ALEPG), Disponível em: <http://www.clul.ulisboa.pt/pt/23-investigacao/681-alepg-atlas-linguistico-etnograficodeportugal-e-da-galiza>. Acesso em: 26 de jan. de 2018.
- AULETE, Caldas. *Novíssimo Aulete dicionário contemporâneo da língua portuguesa*. Rio de Janeiro: Lexikon, 2011.
- AZEREDO, José Carlos de. *Gramática Houaiss da língua portuguesa*. São Paulo: Publifolha, 2008.
- BARBOSA, ALANA BRITO. *A indeterminação do sujeito no falar ludovicense*. 2016. Dissertação (Mestrado em Letras) - Programa de Pós-Graduação em Letras da Universidade Federal do Paraná. Área de Concentração: Estudos Linguísticos. Universidade Federal do Paraná, 2016.
- BARBOSA, Maria Aparecida. *Lexicologia, lexicografia, terminologia, terminografia, identidade científica, objeto, métodos, campos de atuação*. In: Anais do II Simpósio LatinoAmericano de Terminologia. I Encontro Brasileiro de Terminologia Técnico-Científica. Brasília, 1990. p. 152-158.
- BARBOZA, Edson Holanda Lima. *A hidra cearense: rotas de retirantes e escravizados entre o Ceará e as fronteiras do Norte (1877-1884)*. Tese (Doutorado em História Social) - Programa de Pós-Graduação em História, Pontifícia Universidade Católica de São Paulo. São Paulo, 2013. 255f.
- BECHARA, Evanildo. *Moderna gramática portuguesa*. 37 ed. Rio de Janeiro: Lucerna, 2004.
- BORBA, Francisco da Silva. *Dicionário UNESP do português contemporâneo*. São Paulo: UNESP, 2004.
- BORTONI-RICARDO, Stella Maris. *Educação em língua materna: a sociolinguística na sala de aula*. São Paulo: Parábola, 2004.

- CASCUDO, Luís da Câmara. *Dicionário do folclore brasileiro*. Rio de Janeiro: Edições de Ouro, 1969. CASCUDO, Luís da Câmara.
- CARDOSO, Suzana Alice; FERREIRA, Carlota. *A dialetologia no Brasil*. São Paulo: Contexto, 1994.
- CARDOSO, Suzana Alice. *Geolinguística: tradição e modernidade*. São Paulo: Parábola Editorial, 2010.
- CARVALHO, Castelar de. *Para compreender Saussure: fundamentos e visão crítica*.
- Dias, Valter de Carvalho Dias. *A indeterminação do sujeito em textos baianos dos séculos XIX e XX: um estudo sociofuncionalista*. (2017). Tese (Doutorado em Língua e Cultura). Salvador, 2017.
- COSTELLA, Antônio F. *Breve história ilustrada da xilogravura*. Editora Mantiqueira, Campos do Jordão, 2003.
- CAVALCANTE, Vilma Maria Reis. A indeterminação do sujeito no português oral culto de Fortaleza – CE. 1998. 108f. Dissertação (Mestrado em Linguística). Universidade Federal do Ceará, Fortaleza, 1998.
- CEGALLA, Domingos Paschoal. *Novíssima gramática da língua portuguesa*. 48ª ed. São Paulo: Companhia Editora Nacional, (2008 [1977]).
- CEZARIO, Maria Maura; VOTRE, Sebastião. Sociolinguística. In: MARTELOTTA, Mário Eduardo (org.) *Manual de Linguística*. São Paulo: Contexto, 2015.
- COMITÊ NACIONAL DO PROJETO ALiB. *Atlas Linguístico do Brasil*. Questionários. Londrina: Editora da UEL, 2001.
- CUNHA, Celso; CINTRA, Luís F. Lindley. *Gramática da língua portuguesa*. 3 ed. Rio de Janeiro: Nova Fronteira, 2008 [1985].
- CUNHA, Antônio Geraldo da. *Dicionário etimológico da língua portuguesa*. 2.ed. 7.impr. Rio de Janeiro: Nova Fronteira, 1996.
- DOMÍNGUEZ, R. J. *Diccionario nacional de la lengua española*, 14ª ed., Madrid, imprenta y librería de Miguel Guijarro.
- DUARTE, Maria Eugenia. Termos da oração. In: VIEIRA, Silvia Rodrigues; BRANDÃO, Silvia Figueiredo. (org.). *Ensino de gramática: descrição e uso*. São Paulo: Contexto, 2007, p. 185-203.
- RIBEIRO, Silen. No ritmo do bumba meu boi. *FAPEMA*. 2018. Disponível em <https://www.fapema.br/no-ritmo-do-bumba-meu-boi/>. Acesso em 16, abr de 2022.
- FERREIRA, Aurélio Buarque de. *Novo dicionário Aurélio da língua portuguesa*. 3.ed. rev. e atual. Curitiba: Nova Fronteira, 1999
- FEDERAÇÃO DAS INDÚSTRIAS DO ESTADO DO RIO DE JANEIRO. *Índice Firjan de desenvolvimento municipal*. 2016.
- FRANCHI, Carlo. *Mas o que é mesmo “gramática”?*. São Paulo: Parábola Editorial, 2004.
- GARCIA, Carlos. *O que é o Nordeste Brasileiro?*. São Paulo: Editora Brasiliense, 1995.

- GUIMARAES, Tassila Ferreira Vale. *Formas de indeterminação do sujeito no Estado da Bahia: um estudo sociolinguístico*. 2018. Dissertação (Mestrado em Estudo de Linguagens) – Programa de Pós-Graduação em Estudo de Linguagem- PPGEL, Universidade do Estado da Bahia.
- GÂNDAVO, Pero de Magalhães. *História da Província Santa Cruz*. Belo Horizonte, Editora Itatiaia, 1980.
- GUY, Gregory R; ZILLES, Ana. *Sociolinguística quantitativa: instrumental de análise*. São Paulo: Parábola Editorial, 2007.
- HOUAISS, Antônio. *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro, Ed. Objetiva, 2001.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. [2018]. *Censo demográfico população de habitantes*. Disponível em <<https://www.ibge.gov.br/estatisticas/sociais/populacao/17374-indicadores-sociais-minimos.html?=&t=resultados>> Acesso em: 26 set. 2020.
- ILARI, Rodolfo; BASSO, Renato. *O português da gente: a língua que estudamos a língua que falamos*. São Paulo: Contexto, 2017.
- LABOV, William. *Padrões sociolinguísticos*. São Paulo: Parábola, 2008 [1972].
- LÓPEZ MORALES, Humberto. *Sociolingüística*. 3. ed. Madrid: Gredos, 2004.
- MACHADO, José Pedro. *Dicionário etimológico da língua portuguesa*. 2.ed. Lisboa: Confluência, 1990.
- MARTELOTTA, M. E. et alii (org.) 1996. *Gramaticalização no português do Brasil: uma abordagem funcional*. Rio de Janeiro: Tempo Brasileiro/ Departamento de Linguística e Filologia/UFRJ.
- MATEUS, Maria Helena et alii. *Gramática da língua portuguesa*. Lisboa: Editorial Caminho, 2003.
- MELO, Gladstone Chaves de. *A língua do Brasil*. Rio de Janeiro: Padrão, 1981.
- MENON, Odete Pereira da Silva. *Analyse sociolinguistique de l'indétermination du sujets dans le portugais parlé au Brésil, à partir des données du NURC/SP*. 1994. 397f. Tese (Doutorado em Linguística). Université de Paris VII, Paris, [1994] 2006.
- MICHAELIS. *Moderno dicionário da língua portuguesa*. São Paulo: Melhoramentos, 1998. Dicionários Michaelis.
- MILANEZ, Wânia. *Recursos de indeterminação do sujeito*. 1982. 149 f. Dissertação (Mestrado em Linguística) - Universidade Estadual de Campinas, Campinas, 1982.
- MORENO FERNÁNDEZ, Francisco. *Principios de sociolingüística y sociología del lenguaje*. Barcelona: Ariel, 2009.
- NASCENTES, Antenor. *O linguajar carioca*. Rio de Janeiro: Organização Simões, 1922.
- NASCIMENTO, Carina Sampaio, Marcela Moura Torres Paim. "A variação lexical no campo semântico vestuário e acessórios do Projeto Atlas Linguístico do Brasil." *TRAÇOS DE LINGUAGEM-REVISTA DE ESTUDOS LINGÜÍSTICOS* 4.1 (2020). Disponível em

file:///C:/Users/mateu/Downloads/biblioteca,+2+Nascimento+&+Paim.pdf. Acesso em 23 abr. 2022.

NARO, Anthony J. Modelos quantitativos e tratamento estatístico. In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (org.). *Introdução à sociolinguística: o tratamento da variação*. São Paulo: Contexto, 2003. p. 15-25.

NUNES, Odilon. *Pesquisas para a história do Piauí*, v. 2. ed. Rio de Janeiro: Artenova, 1975

OLIVEIRA e SILVA, G. M.; PAIVA, M. C. A. Visão de conjunto das variáveis sociais. In: OLIVEIRA e SILVA, G. M.; SCHERRE, M. M. P. (Orgs.). *Padrões sociolinguísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro*. Rio de Janeiro: Tempo Brasileiro: Departamento de Linguística e Filologia, UFRJ, 1996. cap. 16, p. 335- 378.

PERINI, Mário Alberto. *Gramática descritiva do português*. São Paulo: Editora Ática, 1995.

_____. *Gramática do português brasileiro*. São Paulo: Parábola Editorial, 2010.

PAIVA, Maria da Conceição; SCHERRE, Maria Marta Pereira. “Retrospectiva sociolinguística: contribuições do PEUL”. *DEL TA*, v. 15, n.esp., pp. 201-232, 1999.

PONTES, Eunice S. L. *Sujeito: da sintaxe ao discurso*. São Paulo: Ática; Brasília: INL, Fundação Nacional Pró-Memória. 1986.

PREFEITURA DE SERGIPE. *Aracaju*. Disponível em <<https://www.aracaju.se.gov.br/aracaju/historia/>>. Acesso em: 19 abr. 2022.

RADTKE, E.; THUN, H. Nuevos caminos de la Geolingüística Románica. Un balance. In: *Neue Wege der Romanischen Geolinguistik: Akten des Symposiums zur Empirischen Dialektologie*. Heidelberg/Mainz: Westensee-Verlag Kiel, 1991, p. 25-49.

ROCHA LIMA, Carlos Henrique da. *Gramática normativa da língua portuguesa*. 37 ed. Rio de Janeiro: José Olympio, 2011.

ROMÃO DA SILVA, Júlio. *Memória histórica sobre a transferência da capital do Piauí*. 3. ed. Teresina: Fundação Cultural Monsenhor Chaves, 1994.

RUFFINI, Suseny Maia Teles. “O espaço urbano na literatura de cordel: o olhar de Cuíca de Santo Amaro.” (2009). Dissertação (Mestrado em Geografia), Universidade Federal da Bahia. Disponível em <https://repositorio.ufba.br/bitstream/ri/19370/1/Suseny%20Maia%20Teles%20Ruffini.pdf>. Acesso em 6 abr. 2022.

SANKOFF, David.; TAGLIAMONTE, Sali.; SMITH, Eric. *Goldvarb X: a variable rule application for Macintosh and Windows*. Department of Linguistics, University of Toronto, 2005.

SANTANA, Neila Maria Oliveira. *A indeterminação do sujeito no português rural do semi-árido baiano*. 2006. 153 F. Dissertação (Mestrado em Estudos Linguísticos) – Programa de PósGraduação em Letras e Linguística, Instituto de Letras, Universidade Federal da Bahia. 2012.

SILVA, Giselle Machine de Oliveira; PAIVA, Maria da Conceição. Visão de conjunto das variáveis sociais. In: SILVA, Giselle Machine de Oliveira; SCHERRE, Maria Marta Pereira. Padrões sociolinguísticos. 2 ed. Rio de Janeiro: Tempo Brasileiro, 1998.

SILVA, Ivanilde da. *De quem nós/a gente está(mos) falando afinal?: uma investigação sincrônica da variação entre nós e a gente como estratégias de designação referencial*. 2004. 145 f. Dissertação (Mestrado em Linguística) - Universidade Federal de Santa Catarina.

SILVA, Leonardo Dantas. *Pernambuco: Imagens da Vida e da História*. Recife: SESC, 2001.

TARALLO, Fernando. *A pesquisa sociolinguística*. 2. ed São Paulo: Ática, 2002.

THUN, Harald. La geolinguística como lingüística variacional general (con ejemplos del Atlas lingüístico Diatópico y Diastrático do Uruguay). In: INTERNATIONAL CONGRESS OF ROMANCE LINGUISTICS AND PHILOLOGY. Atti del XXI Congresso internazionale di linguistica e filologia romanza, 21., 1995, Palermo. Tübingen: Niemeyer, 1998, p. 701-729.

VELARDE, Manuel Casado. Las formas fulano, mengano, zutano, perengano y su funcionamiento como ordenadores del discurso. *Lengua y discurso*, Homenaje al Prof. Vidal Lamíquiz, Arco / Libros, Madrid, 2000, pp. 183-195